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. 10, No. 4—December 2016

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. 10, No. 4—December 2016

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1777–1778
DOI: 10.1214/16-AOAS896F
Main article DOI: 10.1214/15-AOAS896

INTRODUCTION TO DISCUSSION OF "COAUTHORSHIP AND CITATION NETWORKS FOR STATISTICIANS"

BY BERNARD W. SILVERMAN

University of Oxford

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1779–1812
DOI: 10.1214/15-AOAS896

COAUTHORSHIP AND CITATION NETWORKS FOR STATISTICIANS

BY PENGSHENG JI AND JIASHUN JIN[2]

University of Georgia and Carnegie Mellon University

We have collected and cleaned two network data sets: Coauthorship and Citation networks for statisticians. The data sets are based on all research papers published in four of the top journals in statistics from 2003 to the first half of 2012. We analyze the data sets from many different perspectives, focusing on (a) productivity, patterns and trends, (b) centrality and (c) community structures.

For (a), we find that over the 10-year period, both the average number of papers per author and the fraction of self citations have been decreasing, but the proportion of distant citations has been increasing. These findings are consistent with the belief that the statistics community has become increasingly more collaborative, competitive and globalized.

For (b), we have identified the most prolific/collaborative/highly cited authors. We have also identified a handful of "hot" papers, suggesting "Variable Selection" as one of the "hot" areas.

For (c), we have identified about 15 meaningful communities or research groups, including large-size ones such as "Spatial Statistics," "Large-Scale Multiple Testing" and "Variable Selection" as well as small-size ones such as "Dimensional Reduction," "Bayes," "Quantile Regression" and "Theoretical Machine Learning."

Our findings shed light on research habits, trends and topological patterns of statisticians. The data sets provide a fertile ground for future research on social networks.

REFERENCES

AMINI, A. A., CHEN, A., BICKEL, P. J. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. MR3127859

ARENAS, A., DUCH, J., FERNÁNDEZ, A. and GÓMEZ, S. (2007). Size reduction of complex networks preserving modularity. *New J. Phys.* **9** 176.1–176.15. MR2335716

BANG-JENSEN, J. and GUTIN, G. (2009). *Digraphs*: *Theory*, *Algorithms and Applications*, 2nd ed. Springer, London. MR2472389

BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

BICKEL, P. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

BICKEL, P. J. and LEVINA, E. (2008a). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

Key words and phrases. Adjacent rand index, centrality, collaboration, community detection, Degree Corrected Block Model, productivity, social network, spectral clustering.

Bickel, P. J. and Levina, E. (2008b). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644

Chen, S. S., Donoho, D. L. and Saunders, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61. MR1639094

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *J. Amer. Statist. Assoc.* **99** 710–723. MR2090905

Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322

Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32** 928–961. MR2065194

Freeman, L. C., Borgatti, S. P. and White, D. R. (1991). Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks* **13** 141–154. MR1135768

Gini, C. (1936). On the measure of concentration with special reference to income and statistics. *Colorado College Publication*, *General Series* **208** 73–79.

Goldenberg, A., Zheng, A., Fienberg, S. and Airoldi, E. (2009). A survey of statistical network models. *Faund. Trends Mach. Learn.* **2** 129–233.

Grossman, J. W. (2002). The evolution of the mathematical research collaboration graph. *Congr. Numer.* **158** 201–212.

Huang, J., Horowitz, J. L. and Ma, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Ann. Statist.* **36** 587–613. MR2396808

Huang, J. Z., Liu, N., Pourahmadi, M. and Liu, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* **93** 85–98. MR2277742

Hubert, L. and Arabie, P. (1985). Comparing partitions. *J. Classification* **2** 193–218.

Hunter, D. R. and Li, R. (2005). Variable selection using MM algorithms. *Ann. Statist.* **33** 1617–1642. MR2166557

Ioannidis, J. P. A. (2008). Measuring co-authorship and networking-adjusted scientific impact. *PLoS ONE* **3** e2778.

Ji, P., Jin, J. and Ke, Z. (2015). Social networks for statisticians, new data and new perspectives. Unpublished manuscript.

Jin, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. MR3285600

Johnstone, I. M. and Silverman, B. W. (2005). Empirical Bayes selection of wavelet thresholds. *Ann. Statist.* **33** 1700–1752. MR2166560

Karrer, B. and Newman, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107, 10. MR2788206

Kim, Y., Son, S.-W. and Jeong, H. (2010). Finding communities in directed networks. *Phys. Rev. E* **81** 016103.

Leicht, E. and Newman, M. (2008). Community structure in directed networks. *Phys. Rev. Lett.* **100** 118703.

Martin, T., Ball, B., Karrer, B. and Newman, M. (2013). Coauthorship and citation patterns in the physical review. *Phys. Rev. E* **88**.

Meila, M. (2003). Comparing clusterings by the variation of information. In *Learning Theory and Kernel Machines*: 16*th Annual Conference on Computational Learning Theory and* 7*th Kernel Workshop* (B. Scholkopf and M. K. Warmuth, eds.). Springer, Berlin.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

Newman, M. E. J. (2001a). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98** 404–409 (electronic). MR1812610

NEWMAN, M. E. J. (2001b). Scientific collaboration networks. I. Network construction and fundamental results. *Phys*. *Rev*. *E* **64** 016131.

NEWMAN, M. (2004). Coauthorship networks and patterns of scientific collaboration. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **101** 5200–5205.

NEWMAN, M. (2006). Modularity and community structure in networks. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **103** 8577–8582.

NEWMAN, M. E. J. and LEICHT, E. A. (2007). Mixture models and exploratory analysis in networks. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **104** 9564–9569.

RAMASCO, J. J. and MUNGAN, M. (2008). Inversion method for content-based networks. *Phys*. *Rev*. *E* (3) **77** 036122, 12. MR2495435

SABIDUSSI, G. (1966). The centrality index of a graph. *Psychometrika* **31** 581–603. MR0205879

STOREY, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann*. *Statist*. **31** 2013–2035. MR2036398

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **58** 267–288. MR1379242

TUKEY, J. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann*. *Statist*. **40** 2266–2292. MR3059083

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J*. *Amer*. *Statist*. *Assoc*. **101** 1418–1429. MR2279469

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **67** 301–320. MR2137327

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann*. *Statist*. **36** 1509–1533. MR2435443

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1813–1819
DOI: 10.1214/16-AOAS976
Main article DOI: 10.1214/15-AOAS896

DISCUSSION OF "COAUTHORSHIP AND CITATION NETWORKS FOR STATISTICIANS"

BY PEDRO REGUEIRO, ABEL RODRÍGUEZ AND JUAN SOSA

University of California

REFERENCES

FRALEY, C. and RAFTERY, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Statist. Assoc.* **97** 611–631. MR1951635

FRALEY, C., RAFTERY, A. E., MURPHY, T. B. and SCRUCCA, L. (2012). mclust Version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation.

GELMAN, A., HWANG, J. and VEHTARI, A. (2014). Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24** 997–1016. MR3253850

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

JI, P. and JIN, J. (2016). Coauthorship and citation networks for statisticians. *Ann. Appl. Stat.* To appear.

KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107, 10. MR2788206

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1820–1826
DOI: 10.1214/16-AOAS977
Main article DOI: 10.1214/15-AOAS896

DISCUSSION OF "COAUTHORSHIP AND CITATION NETWORKS FOR STATISTICIANS"[1]

BY SONG WANG AND KARL ROHE

University of Wisconsin, Madison

Pengsheng Ji and Jiashun Jin have collected and analyzed a fun and fascinating data set that we are eager to use as an example in a course on Statistical Network Analysis. In this comment, we partition the core of the paper citation graph and interpret the clusters by analyzing the paper abstracts using bag-of-words. Under the Stochastic Block Model (SBM), the eigengap reveals the number of clusters. We find several eigengaps and that there are still clusters beyond the largest eigengap. Through this illustration, we argue against a simplistic interpretation of model selection results from the Stochastic Block Model (SBM) literature. In short, don't mind the gap.

REFERENCES

BATES, D. and MAECHLER, M. (2016). Matrix: Sparse and dense matrix classes and methods. R package version 1.2-6. Available at https://CRAN.R-project.org/package=Matrix.

CSARDI, G. and NEPUSZ, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems* **1695**.

JI, P. and JIN, J. (2014). Coauthorship and citation networks for statisticians. Preprint. Available at arXiv:1410.2840.

JIN, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. MR3285600

MEYER, D., HORNIK, K. and FEINERER, I. (2008). Text mining infrastructure in R. *J. Stat. Softw.* **25** 1–54.

QIU, Y. and MEI, J. (2016). rARPACK: Solvers for large scale eigenvalue and svd problems. R package version 0.11-0. Available at https://CRAN.R-project.org/package=rARPACK.

TAI QIN and ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems* 3120–3128.

WANG, S. and ROHE, K. (2016). Supplement to "Discussion of "Coauthorship and citation networks for statisticians"." DOI:10.1214/16-AOAS977SUPP.

Key words and phrases. Networks, spectral clustering, text analysis, eigengap.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1827–1834
DOI: 10.1214/16-AOAS978
Main article DOI: 10.1214/15-AOAS896

DISCUSSION OF "COAUTHORSHIP AND CITATION NETWORKS FOR STATISTICIANS"[1]

BY VISHESH KARWA AND SONJA PETROVIĆ

Harvard University and Illinois Institute of Technology

REFERENCES

BAYARRI, M. J., BERGER, J. O., CAFEO, J., GARCIA-DONATO, G., LIU, F., PALOMO, J., PARTHASARATHY, R. J., PAULO, R., SACKS, J. and WALSH, D. (2007). Computer model validation with functional output. *Ann. Statist.* **35** 1874–1906. MR2363956

BLITZSTEIN, J. and DIACONIS, P. (2010). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Math.* **6** 489–522. MR2809836

DIACONIS, P. and STURMFELS, B. (1998). Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.* **26** 363–397. MR1608156

FIENBERG, S. E., MEYER, M. M. and WASSERMAN, S. S. (1985). Statistical analysis of multiple sociometric relations. *J. Amer. Statist. Assoc.* **80** 51–67.

FIENBERG, S. E. and WASSERMAN, S. S. (1981). Discussion of P. W. Holland and S. Leinhardt "An Exponential Family of Probability Distributions for Directed Graphs". *J. Amer. Statist. Assoc.* **76** 54–57.

GROSS, E., PETROVIĆ, S. and STASI, D. (2015). Goodness-of-fit for log-linear network models: Dynamic Markov bases using hypergraphs. *Ann. Inst. Statist. Math.* DOI: 10.1007/s10463-016-0560-2.

HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176

KARWA, V., PELSMAJER, M., PETROVIĆ, S., STASI, D. and WILBURNE, D. (2016). Statistical models for cores decomposition of an undirected random graph. Submitted.

LUNAGÓMEZ, S., MUKHERJEE, S., WOLPERT, R. L. and AIROLDI, E. M. (2016). Geometric representations of random hypergraphs. *J. Amer. Statist. Assoc.* To appear. Available at: http://www.tandfonline.com/doi/abs/10.1080/01621459.2016.1141686.

PETROVIĆ, S., RINALDO, A. and FIENBERG, S. E. (2010). Algebraic statistics for a directed random graph model with reciprocation. In *Algebraic Methods in Statistics and Probability II* (M. A. G. Viana and H. Wynn, eds.). *Contemporary Mathematics* **516**. Amer. Math. Soc., Providence, RI. MR2605810

RINALDO, A., PETROVIĆ, S. and FIENBERG, S. E. (2013). Maximum likelihood estimation in the β-model. *Ann. Statist.* **41** 1085–1110. MR3113804

SEIDMAN, S. B. (1983). Network structure and minimum degree. *Social Networks* **5** 269–287. MR0721295

STASI, D., SADEGHI, K., RINALDO, A., PETROVIC, S. and FIENBERG, S. (2014). β models for random hypergraphs with a given degree sequence. In *Proceedings of COMPSTAT* 2014—21*st International Conference on Computational Statistics* 593–600. Internat. Statist. Inst., The Hague. MR3372442

ZHU, H., LI, Y., IBRAHIM, J. G., SHI, X., AN, H., CHEN, Y., GAO, W., LIN, W., ROWE, D. B. and PETERSON, B. S. (2009). Regression models for identifying noise sources in magnetic resonance images. *J. Amer. Statist. Assoc.* **104** 623–637. MR2751443

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1835–1841
DOI: 10.1214/16-AOAS896D
Main article DOI: 10.1214/15-AOAS896

DISCUSSION OF "COAUTHORSHIP AND CITATION NETWORKS FOR STATISTICIANS"

BY MLADEN KOLAR[*] AND MATT TADDY[*,†]

University of Chicago[*] and Microsoft Research[†]

REFERENCES

BACALLADO, S. (2011). Bayesian analysis of variable-order, reversible Markov chains. *Ann. Statist.* **39** 838–864. MR2816340

BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

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

BORGATTI, S. P., CARLEY, K. M. and KRACKHARDT, D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Social Networks* **28** 124–136.

BUNEA, F., TSYBAKOV, A. B. and WEGKAMP, M. H. (2007). Aggregation for Gaussian regression. *Ann. Statist.* **35** 1674–1697. MR2351101

CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644

DRTON, M. and PERLMAN, M. D. (2004). Model selection for Gaussian concentration graphs. *Biometrika* **91** 591–602. MR2090624

DRTON, M. and RICHARDSON, T. S. (2004). Multimodality of the likelihood in the bivariate seemingly unrelated regressions model. *Biometrika* **91** 383–392. MR2081308

FAN, J. and LI, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *J. Amer. Statist. Assoc.* **99** 710–723. MR2090905

FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322

FAN, J. and PENG, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32** 928–961. MR2065194

FARCOMENI, A. (2011). Recapture models under equality constraints for the conditional capture probabilities. *Biometrika* **98** 237–242. MR2804224

GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32** 1035–1061. MR2065197

HUANG, J., HOROWITZ, J. L. and MA, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Ann. Statist.* **36** 587–613. MR2396808

HUANG, J. Z., LIU, N., POURAHMADI, M. and LIU, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* **93** 85–98. MR2277742

HUNTER, D. R. and LI, R. (2005). Variable selection using MM algorithms. *Ann. Statist.* **33** 1617–1642. MR2166557

JOHNSTONE, I. M. and SILVERMAN, B. W. (2004). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.* **32** 1594–1649. MR2089135

KOLTCHINSKII, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.* **34** 2593–2656. MR2329442

LEE, Y. K., MAMMEN, E. and PARK, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *Ann. Statist.* **38** 2857–2883. MR2722458

MASSAM, H., LIU, J. and DOBRA, A. (2009). A conjugate prior for discrete hierarchical log-linear models. *Ann. Statist.* **37** 3431–3467. MR2549565

MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

ROBERTS, M. E., STEWART, B. M. and TINGLEY, D. (2014). stm: R package for structural topic models. R package vignette.

ROBERTS, M. E., STEWART, B. M., TINGLEY, D., AIROLDI, E. M. et al. (2013). The structural topic model and applied social science. In *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation*.

SCHICK, A. and WEFELMEYER, W. (2004). Estimating invariant laws of linear processes by U-statistics. *Ann. Statist.* **32** 603–632. MR2060171

STOREY, J. D., TAYLOR, J. E. and SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 187–205. MR2035766

TADDY, M. (2012). On estimation and selection for topic models. In *Proceedings of the* 15*th International Conference on Artificial Intelligence and Statistics* (*AISTATS* 2012).

TADDY, M. (2015). One-step estimator paths for concave regularization. Available at arXiv:1308.5623.

TAN, L. S. L., CHAN, A. H. and ZHENG, T. (2015). Topic-adjusted visibility metric for scientific articles. Available at arXiv:1502.07190.

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. MR2137327

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443

ZOU, H. and ZHANG, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Ann. Statist.* **37** 1733–1751. MR2533470

ZUO, Y. and CUI, H. (2005). Depth weighted scatter estimators. *Ann. Statist.* **33** 381–413. MR2157807

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1842–1845
DOI: 10.1214/16-AOAS896E
Main article DOI: 10.1214/15-AOAS896

DISCUSSION OF "COAUTHORSHIP AND CITATION NETWORKS FOR STATISTICIANS"

BY FORREST W. CRAWFORD

Yale School of Public Health

REFERENCES

BLUNDELL, C., BECK, J. and HELLER, K. A. (2012). Modelling reciprocating relationships with Hawkes processes. In *Advances in Neural Information Processing Systems* 2600–2608.

CHANDRASEKHAR, A. G. and JACKSON, M. O. (2014). Tractable and consistent random graph models. Technical report, National Bureau of Economic Research, Cambridge, MA.

HAWKES, A. G. and OAKES, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Probab.* **11** 493–503. MR0378093

LEE, S. H., KIM, P.-J. and JEONG, H. (2006). Statistical properties of sampled networks. *Phys. Rev. E* (3) **73** 016102.

SHALIZI, C. R. and RINALDO, A. (2013). Consistency under sampling of exponential random graph models. *Ann. Statist.* **41** 508–535. MR3099112

STUMPF, M. P. H., WIUF, C. and MAY, R. M. (2005). Subnets of scale-free networks are not scale-free: Sampling properties of networks. *Proc. Natl. Acad. Sci. USA* **102** 4221–4224.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1846–1853
DOI: 10.1214/16-AOAS896G
Main article DOI: 10.1214/15-AOAS896

REJOINDER: "COAUTHORSHIP AND CITATION NETWORKS FOR STATISTICIANS"

BY PENGSHENG JI AND JIASHUN JIN

University of Georgia and Carnegie Mellon University

REFERENCES

ALBERT, R. and BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys.* **74** 47–97. MR1895096

BICKEL, P. J. and SARKAR, P. (2016). Hypothesis testing for automated community detection in networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 253–273. MR3453655

BONACICH, P. (1972). Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.* **2** 113–120.

DAUDIN, J.-J., PICARD, F. and ROBIN, S. (2008). A mixture model for random graphs. *Stat. Comput.* **18** 173–183. MR2390817

DONOHO, D. (2015). 50 years of data science. *Unpublished manuscript*.

EVANS, J. A. and FOSTER, J. G. (2011). Metaknowledge. *Science* **331** 721–725. MR2798026

GEMAN, D. and GEMAN, S. (2016). Opinion: Science in the age of selfies. *Proc. Natl. Acad. Sci. USA* **113** 9384–9387.

HALL, P. G. (2011). "Ranking our excellence" or "assessing our quality," or whatever. *Inst. Math. Statist. Bull.* **September** 12–14.

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

HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176

IOANNIDIS, J. P. A. (2005). Why most published research findings are false. *PLoS Med.* **2** e124.

JIN, J., KE, Z. T. and LUO, S. (2016). Estimating network memberships by simplex vertices hunting. *Manuscript*.

KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107, 10. MR2788206

KE, Z. T. (2016). A geometrical approach to topic model estimation. Available at arXiv:1608.04478.

LE, C. M. and LEVINA, E. (2015). Estimating the number of communities in networks by spectral methods. Available at arXiv:1507.00827.

NEWMAN, M. (2004). Coauthorship networks and patterns of scientific collaboration. *Proc. Natl. Acad. Sci. USA* **101** 5200–5205.

NEWMAN, M. E. J. (2010). *Networks*: *An Introduction*. Oxford Univ. Press, Oxford. MR2676073

SALDANA, D. F., YU, Y. and FENG, Y. (2016). How many communities are there? *J. Comput. Graph. Statist.* To appear.

STIGLER, S. M. (1994). Citation patterns in the journals of statistics and probability. *Statist. Sci.* **9** 94–108.

STIGLER, G. J., STIGLER, S. M. and FRIEDLAND, C. (1995). The journals of economics. *J. Polit. Econ.* **103** 331–359.

VARIN, C., CATTELAN, M. and FIRTH, D. (2016). Statistical modelling of citation exchange between statistics journals. *J. Roy. Statist. Soc. Ser. A* **179** 1–63.

WANG, D., SONG, C. and BARABÁSI, A.-L. (2013). Quantifying long-term scientific impact. *Science* **342** 127–132.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1854–1879
DOI: 10.1214/16-AOAS975

SMOOTH PRINCIPAL COMPONENT ANALYSIS OVER TWO-DIMENSIONAL MANIFOLDS WITH AN APPLICATION TO NEUROIMAGING

BY EARDI LILA[*,†], JOHN A. D. ASTON[*,1] AND LAURA M. SANGALLI[†]

University of Cambridge[*] *and Politecnico di Milano*[†]

Motivated by the analysis of high-dimensional neuroimaging signals located over the cortical surface, we introduce a novel Principal Component Analysis technique that can handle functional data located over a two-dimensional manifold. For this purpose a regularization approach is adopted, introducing a smoothing penalty coherent with the geodesic distance over the manifold. The model introduced can be applied to any manifold topology, and can naturally handle missing data and functional samples evaluated in different grids of points. We approach the discretization task by means of finite element analysis, and propose an efficient iterative algorithm for its resolution. We compare the performances of the proposed algorithm with other approaches classically adopted in literature. We finally apply the proposed method to resting state functional magnetic resonance imaging data from the Human Connectome Project, where the method shows substantial differential variations between brain regions that were not apparent with other approaches.

REFERENCES

ALFELD, P., NEAMTU, M. and SCHUMAKER, L. L. (1996). Fitting scattered data on sphere-like surfaces using spherical splines. *J. Comput. Appl. Math.* **73** 5–43. MR1424867

BELKIN, M. and NIYOGI, P. (2001). Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems* 14 (T. G. Dietterich, S. Becker and Z. Ghahramani, eds.) 585–591.

BUCKNER, R. L., ANDREWS-HANNA, J. R. and SCHACTER, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* **1124** 1–38.

CAI, D., HE, X., HAN, J. and HUANG, T. S. (2011). Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33** 1548–1560.

CHUNG, M. K., HANSON, J. L. and POLLAK, S. D. (2014). Statistical analysis on brain surfaces. Technical report, University of Wisconsin–Madison.

CHUNG, M. K., ROBBINS, S. M., DALTON, K. M., DAVIDSON, R. J., ALEXANDER, A. L. and EVANS, A. C. (2005). Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage* **25** 1256–1265.

DASSI, F., ETTINGER, B., PEROTTO, S. and SANGALLI, L. M. (2015). A mesh simplification strategy for a spatial regression analysis over the cortical surface of the brain. *Appl. Numer. Math.* **90** 111–131. MR3300898

Key words and phrases. Functional data analysis, principal component analysis, differential regularization, functional magnetic resonance imaging.

DUCHON, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive Theory of Functions of Several Variables* (*Proc. Conf., Math. Res. Inst., Oberwolfach*, 1976) 85–1000. Springer, Berlin. MR0493110

DZIUK, G. (1988). Finite elements for the Beltrami operator on arbitrary surfaces. In *Partial Differential Equations and Calculus of Variations* (S. Hildebrandt and R. Leis, eds.). *Lecture Notes in Math.* **1357** 142–155. Springer, Berlin. MR0976234

ESSEN, D. C. V., UGURBIL, K., AUERBACH, E., BARCH, D., BEHRENS, T. E. J., BUCHOLZ, R., CHANG, A., CHEN, L., CORBETTA, M., CURTISS, S. W., PENNA, S. D., FEINBERG, D., GLASSER, M. F., HAREL, N., HEATH, A. C., LARSON-PRIOR, L., MARCUS, D., MICHALAREAS, G., MOELLER, S., OOSTENVELD, R., PETERSEN, S. E., PRIOR, F., SCHLAGGAR, B. L., SMITH, S. M., SNYDER, A. Z., XU, J. and YACOUB, E. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage* **62** 2222–2231.

ETTINGER, B., PEROTTO, S. and SANGALLI, L. M. (2016). Spatial regression models over two-dimensional manifolds. *Biometrika* **103** 71–88. MR3465822

GLASSER, M. F., SOTIROPOULOS, S. N., WILSON, J. A., COALSON, T. S., FISCHL, B., ANDERSSON, J. L., XU, J., JBABDI, S., WEBSTER, M., POLIMENI, J. R., ESSEN, D. C. V. and JENKINSON, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80** 105–124.

GORDON, E. M., LAUMANN, T. O., ADEYEMO, B., HUCKINS, J. F., KELLEY, W. M. and PETERSEN, S. E. (2014). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex*.

GREEN, P. J. and SILVERMAN, B. W. (1993). *Nonparametric Regression and Generalized Linear Models*. CRC Press, Boca Raton.

HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 109–126. MR2212577

HARRISON, S. J., WOOLRICH, M. W., ROBINSON, E. C., GLASSER, M. F., BECKMANN, C. F., JENKINSON, M. and SMITH, S. M. (2015). Large-scale probabilistic functional modes from resting state fMRI. *NeuroImage* **109** 217–231.

HUANG, J. Z., SHEN, H. and BUJA, A. (2008). Functional principal components analysis via penalized rank one approximation. *Electron. J. Stat.* **2** 678–695. MR2426107

JOLLIFFE, I. T., TRENDAFILOV, N. T. and UDDIN, M. (2003). A modified principal component technique based on the LASSO. *J. Comput. Graph. Statist.* **12** 531–547. MR2002634

LILA, E., ASTON, J. A. D. and SANGALLI, L. M. Supplement to “Smooth Principal Component Analysis over two-dimensional manifolds with an application to neuroimaging.” DOI:10.1214/16-AOAS975SUPP.

MARRON, J. S., RAMSAY, J. O., SANGALLI, L. M. and SRIVASTAVA, A. (2015). Functional data analysis of amplitude and phase variation. *Statist. Sci.* **30** 468–484. MR3432837

OGAWA, S., LEE, T. M., KAY, A. R. and TANK, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA* **87** 9868–9872.

RAMSAY, T. (2002). Spline smoothing over difficult regions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 307–319. MR1904707

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

RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** 233–243. MR1094283

RIESZ, F. and SZ.-NAGY, B. (1955). *Functional Analysis*. Frederick Ungar Publishing Co., New York. MR0071727

SANGALLI, L. M., RAMSAY, J. O. and RAMSAY, T. O. (2013). Spatial spline regression models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 681–703. MR3091654

SHEN, H. and HUANG, J. Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *J. Multivariate Anal.* **99** 1015–1034. MR2419336

SILVERMAN, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.* **24** 1–24. MR1389877

WAHBA, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Statist. Comput.* **2** 5–16. MR0618629

ZHOU, L. and PAN, H. (2014). Principal component analysis of two-dimensional functional data. *J. Comput. Graph. Statist.* **23** 779–801. MR3224656

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 301–320. MR2137327

ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2006). Sparse principal component analysis. *J. Comput. Graph. Statist.* **15** 265–286. MR2252527

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1880–1906
DOI: 10.1214/16-AOAS947

LINKING LUNG AIRWAY STRUCTURE TO PULMONARY FUNCTION VIA COMPOSITE BRIDGE REGRESSION

BY KUN CHEN[1,2,*], ERIC A. HOFFMAN[1,†], INDU SEETHARAMAN[‡], FEIRAN JIAO[†], CHING-LONG LIN[1,†] AND KUNG-SIK CHAN[1,†]

University of Connecticut,[*] University of Iowa[†] and Kansas State University [‡]

The human lung airway is a complex inverted tree-like structure. Detailed airway measurements can be extracted from MDCT-scanned lung images, such as segmental wall thickness, airway diameter, parent-child branch angles, etc. The wealth of lung airway data provides a unique opportunity for advancing our understanding of the fundamental structure-function relationships within the lung. An important problem is to construct and identify important lung airway features in normal subjects and connect these to standardized pulmonary function test results such as FEV1%. Among other things, the problem is complicated by the fact that a particular airway feature may be an important (relevant) predictor only when it pertains to segments of certain generations. Thus, the key is an efficient, consistent method for simultaneously conducting group selection (lung airway feature types) and within-group variable selection (airway generations), i.e., bi-level selection. Here we streamline a comprehensive procedure to process the lung airway data via imputation, normalization, transformation and groupwise principal component analysis, and then adopt a new composite penalized regression approach for conducting bi-level feature selection. As a prototype of composite penalization, the proposed composite bridge regression method is shown to admit an efficient algorithm, enjoy bi-level oracle properties and outperform several existing methods. We analyze the MDCT lung image data from a cohort of 132 subjects with normal lung function. Our results show that lung function in terms of FEV1% is promoted by having a less dense and more homogeneous lung comprising an airway whose segments enjoy more heterogeneity in wall thicknesses, larger mean diameters, lumen areas and branch angles. These data hold the potential of defining more accurately the "normal" subject population with borderline atypical lung functions that are clearly influenced by many genetic and environmental factors.

REFERENCES

BECKLAKE, M. R. (1985). Concepts of normality applied to the measurement of lung function. *Am. J. Med.* **80** 1158–1164.

BREHENY, P. (2015). The group exponential lasso for bi-level variable selection. *Biometrics* **71** 731–740. MR3402609

BREHENY, P. and HUANG, J. (2009). Penalized methods for bi-level variable selection. *Stat. Interface* **2** 369–380. MR2540094

Key words and phrases. Bi-level variable selection, composite penalization, feature extraction, lung airway data, pulmonary function tests.

BREHENY, P. and HUANG, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat.* **5** 232–253. MR2810396

CHEN, K. and CHAN, K.-S. (2011). Subset ARMA selection via the adaptive Lasso. *Stat. Interface* **4** 197–205. MR2812815

CHEN, K., CHAN, K.-S. and STENSETH, N. CHR. (2012). Reduced rank stochastic regression with a sparse singular value decomposition. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 203–221. MR2899860

CHEN, L. and HUANG, J. Z. (2012). Sparse reduced-rank regression for simultaneous dimension reduction and variable selection. *J. Amer. Statist. Assoc.* **107** 1533–1545. MR3036414

CHEN, K., HOFFMAN, E. A., SEETHARAMAN, I., JIAO, F., LIN, C.-L. and CHAN, K.-S. (2016). Supplement to "Linking lung airway structure to pulmonary function via composite bridge regression." DOI:10.1214/16-AOAS947SUPP.

EFRON, B. (2004). The estimation of prediction error: Covariance penalties and cross-validation. *J. Amer. Statist. Assoc.* **99** 619–642. MR2090899

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

FRIEDMAN, J. H., HASTIE, T. J. and TIBSHIRANI, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22.

FULD, M. K., GROUT, R. W., GUO, J., MORGAN, J. H. and HOFFMAN, E. (2012). Systems for lung volume standardization during static and dynamic MDCT-based quantitative assessment of pulmonary structure and function. *Acad. Radiol.* **19** 930–940.

GAO, W. (2010). Development of human lung query atlas. Dissertation, Univ. Iowa.

GUO, J., FULD, M. K., ALFORD, S. K., REINHARDT, J. M. and HOFFMAN, E. A. (2008). Pulmonary Analysis Software Suite 9.0: Integrating quantitative measures of function with structural analyses. In *First International Workshop on Pulmonary Image Analysis* 283–292.

HANKINSON, J. L., ODENCRANTZ, J. R. and FEDAN, K. B. (1999). Spirometric reference values from a sample of the general U.S. population. *Am. J. Respir. Crit. Care Med.* **159** 179–187.

HOFFMAN, E. A., SIMON, B. A. and MCLENNAN, G. (2006). State of the art. A structural and functional assessment of the lung via multidetector-row computed tomography: Phenotyping chronic obstructive pulmonary disease. *Proc. Am. Thorac. Soc.* **3** 519–532.

HUANG, J., BREHENY, P. and MA, S. (2012). A selective review of group selection in high-dimensional models. *Statist. Sci.* **27** 481–499. MR3025130

HUANG, J., HOROWITZ, J. L. and MA, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Ann. Statist.* **36** 587–613. MR2396808

HUANG, J., MA, S., XIE, H. and ZHANG, C.-H. (2009). A group bridge approach for variable selection. *Biometrika* **96** 339–355. MR2507147

IYER, K. S., GRANT, R. W., ZAMBA, G. K. and HOFFMAN, E. A. (2014). Repeatability and sample size assessment associated with computed tomography-based lung density metrics. *Journal of the COPD Foundation* **1** 97–104.

LIU, J., MA, S. and HUANG, J. (2014). Integrative analysis of cancer diagnosis studies with composite penalization. *Scand. Stat. Theory Appl.* **41** 87–103. MR3181134

MA, S., HUANG, J., WEI, F., XIE, Y. and FANG, K. (2011). Integrative analysis of multiple cancer prognosis studies with gene expression measurements. *Stat. Med.* **30** 3361–3371. MR2861619

MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

MONTESANTOS, S., KATZ, I., FLEMING, J., MAJORAL, C., PICHELIN, M., DUBAU, C., PIEDNOIR, B., CONWAY, J., TEXEREAU, J. and CAILLIBOTTE, G. (2013). Airway morphology from high resolution computed tomography in healthy subjects and patients with moderate persistent asthma. *Anat Rec* (*Hoboken*) **296** 852–866.

NAKANO, Y., THO, N. V., YAMADA, H., OSAWA, M. and NAGAO, T. (2009). Radiological approach to asthma and COPD—the role of computed tomography. *Allergol. Intern.* **58** 323–331.

PALAGYI, K., TSCHIRREN, J., HOFFMAN, E. A. and SONKA, M. (2006). Quantitative analysis of pulmonary airway tree structuress. *Comput. Biol. Med.* **36** 974–976.

R DEVELOPMENT CORE TEAM (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

SMITH, B. M., HOFFMAN, E. A., RABINOWITZ, D., BLEECKER, E., CHRISTENSON, S., COUPER, D., DONOHUE, K. M., HAN, M. K., HANSEL, N. N., KANNER, R. E. et al. (2014). Comparison of spatially matched airways reveals thinner airway walls in COPD. The Multi-Ethnic Study of Atherosclerosis (MESA) COPD Study and the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). *Thorax* **69** 987–996.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

TSCHIRREN, J., HOFFMAN, E. A., MCLENNAN, G. and SONKA, M. (2005a). Intrathoracic airway trees: Segmentation and airway morphology analysis from low-dose CT scans. *IEEE Trans. Med. Imag.* **24** 1529–1539.

TSCHIRREN, J., HOFFMAN, E. A., MCLENNAN, G. and SONKA, M. (2005b). Segmentation and quantitative analysis of intrathoracic airway trees from computed tomography images. *Proc. Am. Thorac. Soc.* **2** 484–7, 503–4.

TSCHIRREN, J., MCLENNAN, G., PALAGYI, K., HOFFMAN, E. A. and SONKA, M. (2005c). Matching and anatomical labeling of human airway tree. *Comput. Biol. Med.* **24** 1540–1547.

WEIBEL, E. R. (2015). How Benoit Mandelbrot changed my thinking about biological form. *Benoit Mandelbrot: A Life in Many Dimensions* **1** 471–487.

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

ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701

ZHANG, C., JIANG, Y. and CHAI, Y. (2010). Penalized Bregman divergence for large-dimensional regression and classification. *Biometrika* **97** 551–566. MR2672483

ZHAO, P., ROCHA, G. and YU, B. (2009). The composite absolute penalties family for grouped and hierarchical variable selection. *Ann. Statist.* **37** 3468–3497. MR2549566

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1907–1929
DOI: 10.1214/16-AOAS925

CATEGORICAL DATA FUSION USING AUXILIARY INFORMATION[1]

BY BAILEY K. FOSDICK*, MARIA DEYOREO† AND JEROME P. REITER†

Colorado State University and Duke University†*

In data fusion, analysts seek to combine information from two databases comprised of disjoint sets of individuals, in which some variables appear in both databases and other variables appear in only one database. Most data fusion techniques rely on variants of conditional independence assumptions. When inappropriate, these assumptions can result in unreliable inferences. We propose a data fusion technique that allows analysts to easily incorporate auxiliary information on the dependence structure of variables not observed jointly; we refer to this auxiliary information as glue. With this technique, we fuse two marketing surveys from the book publisher HarperCollins using glue from the online, rapid-response polling company CivicScience. The fused data enable estimation of associations between people's preferences for authors and for learning about new books. The analysis also serves as a case study on the potential for using online surveys to aid data fusion.

REFERENCES

D'ORAZIO, M., DI ZIO, M. and SCANU, M. (2006). *Statistical Matching*: *Theory and Practice*. Wiley, Chichester. MR2268833

D'ORAZIO, M., DI ZIO, M. and SCANU, M. (2002). Statistical matching and official statistics. *Rivista di Statistica Ufficiale* **1** 5–24.

DUNSON, D. B. and XING, C. (2009). Nonparametric Bayes modeling of multivariate categorical data. *J. Amer. Statist. Assoc.* **104** 1042–1051. MR2562004

FOSDICK, B., DEYOREO, M. and REITER, J. (2016). Supplement to "Categorical data fusion using auxiliary information." DOI:10.1214/16-AOAS925SUPP.

GIBBS, A. and SU, F. (2002). On choosing and bounding probability metrics. *Int. Stat. Rev.* **70** 419–435.

GILULA, Z. and MCCULLOCH, R. (2013). Multi level categorical data fusion using partially fused data. *Quantitative Marketing and Economics* **11** 353–377.

GILULA, Z., MCCULLOCH, R. and ROSSI, P. (2006). A direct approach to data fusion. *Journal of Marketing Research* **43** 73–83.

GOODMAN, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61** 215–231. MR0370936

ISHWARAN, H. and JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96** 161–173. MR1952729

ISHWARAN, H. and ZAREPOUR, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* **87** 371–390. MR1782485

KADANE, J. B. (2001). Some statistical problems in merging data files. *Journal of Official Statistics* **17** 423–433.

Key words and phrases. Imputation, integration, latent class, matching.

KAMAKURA, W. and WEDEL, M. (1997). Statistical data fusion for cross tabulation. *Journal of Marketing Research* **34** 485–498.

KAMAKURA, W., WEDEL, M., DE ROSA, F. and MAZZON, J. A. (2003). Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction. *International Journal of Research in Marketing* **20** 45–65.

KIESL, H. and RÄSSLER, S. (2006). How valid can data fusion be? IAB Discussion Paper, 15.

MORIARITY, C. and SCHEUREN, F. (2003). A note on Rubin's statistical matching using file concatenation with adjusted weights and multiple imputations. *J. Bus. Econom. Statist.* **21** 65–73. MR1973805

MORIARTY, C. and SCHEUREN, F. (2001). Statistical matching: A paradigm for assessing the uncertainty in the procedure. *Journal of Official Statistics* **17** 407–422.

POLLARD (2002). *A User's Guide to Measure Theoretic Probability*. Cambridge Univ. Press, Cambridge.

RÄSSLER, S. (2002). *Statistical Matching*: *A Frequentist Theory*, *Practical Applications*, *and Alternative Bayesian Approaches*. *Lecture Notes in Statistics* **168** 60–63. Springer, New York. MR1996879

RÄSSLER, S. (2004). Data fusion: Identification problems, validity, and multiple imputation. *Austrian Journal of Statistics* **33** 153–171.

REITER, J. P. (2012). Bayesian finite population imputation for data fusion. *Statist. Sinica* **22** 795–811. MR2954362

RODGERS, W. L. (1994). An evaluation of statistical matching. *J. Bus. Econom. Statist.* **2** 91–102.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196

RUBIN, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *J. Bus. Econom. Statist.* **4** 87–94.

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York. MR0899519

SCHIFELING, T. A. and REITER, J. P. (2016). Incorporating marginal prior information in latent class models. *Bayesian Anal.* **11** 499–518. MR3472000

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

SI, Y. and REITER, J. P. (2013). Nonparametric Bayesian multiple imputation for incomplete categorical variables in large-scale assessment surveys. *Journal of Educational and Behavioral Statistics* **38** 499–521.

VAN HATTUM, P. and HOIJTINK, H. (2008). The proof of the pudding is in the eating. Data fusion: An application in marketing. *Journal of Database Marketing & Customer Strategy Management* **15** 267–284.

VAN DER PUTTEN, P., KOK, J. N. and GUPTA, A. (2002). Data fusion through statistical matching. Working paper 4342-02. MIT Sloan School of Management, Cambridge, MA.

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

WICKEN, G. and ELMS, S. (2009). Demystifying data fusion—The "why?", the "how?" and the "wow!" Technical report, Advertising Research Foundation Week of Workshops, New York.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1930–1957
DOI: 10.1214/16-AOAS946
© Institute of Mathematical Statistics, 2016

INVESTIGATING DIFFERENCES IN BRAIN FUNCTIONAL NETWORKS USING HIERARCHICAL COVARIATE-ADJUSTED INDEPENDENT COMPONENT ANALYSIS[1]

BY RAN SHI AND YING GUO

Emory University

Human brains perform tasks via complex functional networks consisting of separated brain regions. A popular approach to characterize brain functional networks in fMRI studies is independent component analysis (ICA), which is a powerful method to reconstruct latent source signals from their linear mixtures. In many fMRI studies, an important goal is to investigate how brain functional networks change according to specific clinical and demographic variabilities. Existing ICA methods, however, cannot directly incorporate covariate effects in ICA decomposition. Heuristic post-ICA analysis to address this need can be inaccurate and inefficient. In this paper, we propose a hierarchical covariate-adjusted ICA (hc-ICA) model that provides a formal statistical framework for estimating covariate effects and testing differences between brain functional networks. Our method provides a more reliable and powerful statistical tool for evaluating group differences in brain functional networks while appropriately controlling for potential confounding factors. We present an analytically tractable EM algorithm to obtain maximum likelihood estimates of our model. We also develop a subspace-based approximate EM that runs significantly faster while retaining high accuracy. To test the differences in functional networks, we introduce a voxel-wise approximate inference procedure which eliminates the need of computationally expensive covariance matrix estimation and inversion. We demonstrate the advantages of our methods over the existing method via simulation studies. We apply our method to an fMRI study to investigate differences in brain functional networks associated with post-traumatic stress disorder (PTSD).

REFERENCES

ANAND, A., LI, Y., WANG, Y., WU, J., GAO, S., BUKHARI, L., MATHEWS, V. P., KALNIN, A. and LOWE, M. J. (2005). Activity and connectivity of brain mood regulating circuit in depression: A functional magnetic resonance study. *Biol. Psychiatry* **57** 1079–1088.

ATTIAS, H. (1999). Independent factor analysis. *Neural Comput.* **11** 803–851.

ATTIAS, H. (2000). A variational Bayesian framework for graphical models. *Adv. Neural Inf. Process. Syst.* **12** 209–215.

BECK, A. T., STEER, R. A. and CARBIN, M. G. (1988). Psychometric properties of the beck depression inventory: Twenty-five years of evaluation. *Clin. Psychol. Rev.* **8** 77–100.

BECK, A. T., STEER, R. A., BROWN, G. K. et al. (1996). Manual for the beck depression inventory-II.

Key words and phrases. fMRI, blind source separation, brain functional networks, EM algorithm, subspace concentration.

BECKMANN, C. F. and SMITH, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imag.* **23** 137–152.

BECKMANN, C. F. and SMITH, S. M. (2005). Tensorial extensions of independent component analysis for multisubject FMRI analysis. *NeuroImage* **25** 294–311.

BECKMANN, C. F., DELUCA, M., DEVLIN, J. T. and SMITH, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **360** 1001–1013.

BECKMANN, C. F., MACKAY, C. E., FILIPPINI, N. and SMITH, S. M. (2009). Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. *NeuroImage* **47** S148.

BISWAL, B. B. and ULMER, J. L. (1999). Blind source separation of multiple signal sources of fMRI data sets using independent component analysis. *J. Comput. Assist. Tomogr.* **23** 265–271.

BULLMORE, E. and SPORNS, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev., Neurosci.* **10** 186–198.

BULLMORE, E., BRAMMER, M., WILLIAMS, S. C., RABE-HESKETH, S., JANOT, N., DAVID, A., MELLERS, J., HOWARD, R. and SHAM, P. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magn. Reson. Med.* **35** 261–277.

CALHOUN, V. D., ADALI, T., PEARLSON, G. D. and PEKAR, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* **14** 140–151.

CAMPBELL, D. G., FELKER, B. L., LIU, C.-F., YANO, E. M., KIRCHNER, J. E., CHAN, D., RUBENSTEIN, L. V. and CHANEY, E. F. (2007). Prevalence of depression–PTSD comorbidity: Implications for clinical practice guidelines and primary care-based interventions. *Journal of General Internal Medicine* **22** 711–718.

CHEN, C.-H., RIDLER, K., SUCKLING, J., WILLIAMS, S., FU, C. H., MERLO-PICH, E. and BULLMORE, E. (2007). Brain imaging correlates of depressive symptom severity and predictors of symptom improvement after antidepressant treatment. *Biol. Psychiatry* **62** 407–414.

COLE, L. J., FARRELL, M. J., GIBSON, S. J. and EGAN, G. F. (2010). Age-related differences in pain sensitivity and regional brain activity evoked by noxious pressure. *Neurobiol. Aging* **31** 494–503.

DANIELS, J. K., FREWEN, P., MCKINNON, M. C. and LANIUS, R. A. (2011). Default mode alterations in posttraumatic stress disorder related to early-life trauma: A developmental perspective. *J. Psychiatry Neurosci.* **36** 56–59.

DAUBECHIES, I., ROUSSOS, E., TAKERKART, S., BENHARROSH, M., GOLDEN, C., D'ARDENNE, K., RICHTER, W., COHEN, J. D. and HAXBY, J. (2009). Independent component analysis for brain fMRI does not select for independence. *Proc. Natl. Acad. Sci. USA* **106** 10415–10422.

FILIPPINI, N., MACINTOSH, B. J., HOUGH, M. G., GOODWIN, G. M., FRISONI, G. B., SMITH, S. M., MATTHEWS, P. M., BECKMANN, C. F. and MACKAY, C. E. (2009). Distinct patterns of brain activity in young carriers of the APOE-$\varepsilon4$ allele. *Proc. Natl. Acad. Sci. USA* **106** 7209–7214.

FIRST, M. B. (1995). Structured clinical interview for the DSM (SCID). In *The Encyclopedia of Clinical Psychology*. American Psychiatric Press, Washington, DC.

GENOVESE, C. R., LAZAR, N. A. and NICHOLS, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* **15** 870–878.

GREICIUS, M. D., FLORES, B. H., MENON, V., GLOVER, G. H., SOLVASON, H. B., KENNA, H., REISS, A. L. and SCHATZBERG, A. F. (2007). Resting-state functional connectivity in major depression: Abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biol. Psychiatry* **62** 429–437.

GRIFFANTI, L., SALIMI-KHORSHIDI, G., BECKMANN, C. F., AUERBACH, E. J., DOUAUD, G., SEXTON, C. E., ZSOLDOS, E., EBMEIER, K. P., FILIPPINI, N., MACKAY, C. E. et al. (2014).

ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage* **95** 232–247.

GUO, Y. (2011). A general probabilistic model for group independent component analysis and its estimation methods. *Biometrics* **67** 1532–1542. MR2872404

GUO, Y. and PAGNONI, G. (2008). A unified framework for group independent component analysis for multi-subject fMRI data. *NeuroImage* **42** 1078–1093.

GUO, Y. and TANG, L. (2013). A hierarchical model for probabilistic independent component analysis of multi-subject fMRI studies. *Biometrics* **69** 970–981. MR3146792

HENDLER, T., ROTSHTEIN, P., YESHURUN, Y., WEIZMANN, T., KAHN, I., BEN-BASHAT, D., MALACH, R. and BLEICH, A. (2003). Sensing the invisible: Differential sensitivity of visual cortex and amygdala to traumatic context. *NeuroImage* **19** 587–600.

HIMBERG, J., HYVÄRINEN, A. and ESPOSITO, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage* **22** 1214–1222.

HYVÄRINEN, A., KARHUNEN, J. and OJA, E. (2001). *Independent Component Analysis* **46**. Wiley, New York.

HYVÄRINEN, A. and OJA, E. (2000). Independent component analysis: Algorithms and applications. *Neural Netw.* **13** 411–430.

KESSLER, R. C., SONNEGA, A., BROMET, E., HUGHES, M. and NELSON, C. B. (1995). Posttraumatic stress disorder in the national comorbidity survey. *Arch. Gen. Psychiatry* **52** 1048–1060.

KOSTANTINOS, N. (2000). Gaussian mixtures and their applications to signal processing. In *Advanced Signal Processing Handbook*. CRC Press, New York.

LEE, S., SHEN, H., TRUONG, Y., LEWIS, M. and HUANG, X. (2011). Independent component analysis involving autocorrelated sources with an application to functional magnetic resonance imaging. *J. Amer. Statist. Assoc.* **106** 1009–1024. MR2894760

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

MCKEOWN, M. J., MAKEIG, S., BROWN, G. G., JUNG, T.-P., KINDERMANN, S. S., KINDERMANN, R. S., BELL, A. J. and SEJNOWSKI, T. J. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* **6** 160–188.

MCLACHLAN, G. and PEEL, D. (2004). *Finite Mixture Models*. Wiley, New York.

MEILIJSON, I. (1989). A fast improvement to the EM algorithm on its own terms. *J. Roy. Statist. Soc. Ser. B* **51** 127–138. MR0984999

MENG, X.-L. and RUBIN, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *J. Amer. Statist. Assoc.* **86** 899–909.

MINKA, T. P. (2000). Automatic choice of dimensionality for PCA. In *NIPS* **13** 598–604. MIT Press, Cambridge, MA.

QUITON, R. L. and GREENSPAN, J. D. (2007). Sex differences in endogenous pain modulation by distracting and painful conditioning stimulation. *Pain* **132 Suppl 1** S134–S149.

RAICHLE, M. E., MACLEOD, A. M., SNYDER, A. Z., POWERS, W. J., GUSNARD, D. A. and SHULMAN, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. USA* **98** 676–682.

REINEBERG, A. E., ANDREWS-HANNA, J. R., DEPUE, B. E., FRIEDMAN, N. P. and BANICH, M. T. (2015). Resting-state networks predict individual differences in common and specific aspects of executive function. *NeuroImage* **104** 69–78.

SEBER, G. A. and LEE, A. J. (2012). *Linear Regression Analysis* **936**. Wiley, New York.

SHELINE, Y. I., BARCH, D. M., PRICE, J. L., RUNDLE, M. M., VAISHNAVI, S. N., SNYDER, A. Z., MINTUN, M. A., WANG, S., COALSON, R. S. and RAICHLE, M. E. (2009). The default mode network and self-referential processes in depression. *Proc. Natl. Acad. Sci. USA* **106** 1942–1947.

SHI, R. and GUO, Y. (2016). Supplement to "Investigating differences in brain functional networks using hierarchical covariate-adjusted independent component analysis." DOI:10.1214/16-AOAS946SUPP.

SMITH, S. M., FOX, P. T., MILLER, K. L., GLAHN, D. C., FOX, P. M., MACKAY, C. E., FILIPPINI, N., WATKINS, K. E., TORO, R., LAIRD, A. R. et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. USA* **106** 13040–13045.

SMITH, D. V., UTEVSKY, A. V., BLAND, A. R., CLEMENT, N., CLITHERO, J. A., HARSCH, A. E., CARTER, R. M. and HUETTEL, S. A. (2014). Characterizing individual differences in functional connectivity using dual-regression and seed-based approaches. *NeuroImage* **95** 1–12.

TOHKA, J., FOERDE, K., ARON, A. R., TOM, S. M., TOGA, A. W. and POLDRACK, R. A. (2008). Automatic independent component labeling for artifact removal in fMRI. *NeuroImage* **39** 1227–1245.

WHITFIELD-GABRIELI, S., THERMENOS, H. W., MILANOVIC, S., TSUANG, M. T., FARAONE, S. V., MCCARLEY, R. W., SHENTON, M. E., GREEN, A. I., NIETO-CASTANON, A., LAVIOLETTE, P. et al. (2009). Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proc. Natl. Acad. Sci. USA* **106** 1279–1284.

XU, L., CHEUNG, C., YANG, H. and AMARI, S. (1997). Maximum equalization by entropy maximization and mixture of cumulative distribution functions. In *Proc. of ICNN'97* 1821–1826. Springer, New York.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1958–1976
DOI: 10.1214/16-AOAS959

IMPROVING COVARIATE BALANCE IN 2^K FACTORIAL DESIGNS VIA RERANDOMIZATION WITH AN APPLICATION TO A NEW YORK CITY DEPARTMENT OF EDUCATION HIGH SCHOOL STUDY

BY ZACH BRANSON, TIRTHANKAR DASGUPTA AND DONALD B. RUBIN

Harvard University

A few years ago, the New York Department of Education (NYDE) was planning to conduct an experiment involving five new intervention programs for a selected set of New York City high schools. The goal was to estimate the causal effects of these programs and their interactions on the schools' performance. For each of the schools, about 50 premeasured covariates were available. The schools could be randomly assigned to the 32 treatment combinations of this 2^5 factorial experiment, but such an allocation could have resulted in a huge covariate imbalance across treatment groups. Standard methods used to prevent confounding of treatment effects with covariate effects (e.g., blocking) were not intuitive due to the large number of covariates. In this paper, we explore how the recently proposed and studied method of rerandomization can be applied to this problem and other factorial experiments. We propose how to implement rerandomization in factorial experiments, extend the theoretical properties of rerandomization from single-factor experiments to 2^K factorial designs, and demonstrate, using the NYDE data, how such a designed experiment can improve precision of estimated factorial effects.

REFERENCES

AHLUWALIA, J. S., OKUYEMI, K., NOLLEN, N., CHOI, W. S., KAUR, H., PULVERS, K. and MAYO, M. S. (2006). The effects of nicotine gum and counseling among African American light smokers: A 2 × 2 factorial design. *Addiction* **101** 883–891.

APFEL, C. C., KRANKE, P., KATZ, M. H., GOEPFERT, C., PAPENFUSS, S., RAUCH, S., HEINECK, R., GREIM, C. A. and ROEWER, R. (2002). Volatile anaesthetics may be the main cause of early but not delayed postoperative vomiting: A randomized controlled trial of factorial design. *Br. J. Anaesth.* **88** 659–668.

BAYS, H. E., OSE, L., FRASER, N., TRIBBLE, D. L., QUINTO, K., REYES, R., JOHNSON-LEVONAS, A. O., SAPRE, A., DONAHUE, S. R. and EZETIMIBE STUDY GROUP (2004). A multicenter, randomized, double-blind, placebo-controlled, factorial design study to evaluate the lipid-altering efficacy and safety profile of the ezetimibe/simvastatin tablet compared with ezetimibe and simvastatin monotherapy in patients with primary hypercholesterolemia. *Clin. Ther.* **26** 1758–1773.

BOX, G. E. P., HUNTER, J. S. and HUNTER, W. G. (2005). *Statistics for Experimenters*: *Design*, *Innovation*, *and Discovery*, 2nd ed. Wiley, Hoboken, NJ. MR2140250

Key words and phrases. Experimental design, treatment allocation, randomization, Mahalanobis distance, factorial effects.

BRANSON, Z., DASGUPTA, T. and RUBIN, D. B. (2016). Supplement to "Improving covariate balance in 2^K factorial designs via rerandomization with an application to a New York City Department of Education High School Study." DOI:10.1214/16-AOAS959SUPP.

BRUHN, M. and MCKENZIE, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *Am. Econ. J. Appl. Econ.* **1** 200–232.

COX, D. R. (2009). Randomization in the design of experiments. *Int. Stat. Rev.* **77** 415–429.

DASGUPTA, T., PILLAI, N. S. and RUBIN, D. B. (2015). Causal inference from 2^K factorial designs by using potential outcomes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 727–753. MR3382595

ESPINOSA, V., DASGUPTA, T. and RUBIN, D. B. (2016). A Bayesian perspective on the analysis of unreplicated factorial experiments using potential outcomes. *Technometrics* **58** 62–73. MR3463157

FISHER, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

FISHER, R. A. (1942). *The Design of Experiments*, 3rd ed. ed. Hafner-Publishing, New York.

GU, X. S. and ROSENBAUM, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *J. Comput. Graph. Statist.* **2** 405–420.

HU, Y. and HU, F. (2012). Asymptotic properties of covariate-adaptive randomization. *Ann. Statist.* **40** 1794–1815. MR3015044

KASARI, C., ROTHERAM-FULLER, E., LOCKE, J. and GULSRUD, A. (2012). Making the connection: Randomized controlled trial of social skills at school for children with autism spectrum disorders. *J. Child Psychol. Psychiatry* **53** 431–439.

KOLLAR, I., FISCHER, F. and SLOTTA, J. D. (2005). Internal and external collaboration scripts in web-based science learning at schools. In *Proceedings of the* 2005 *Conference on Computer Support for Collaborative Learning*: *Learning* 2005: *The Next* 10 *Years! CSCL* '05, *Taipei*, *Taiwan*, *May* 30–*June* 4, 2005. 331–340. International Society of the Learning Sciences.

KRAUSE, M. S. and HOWARD, K. I. (2003). What random assignment does and does not do. *Journal of Clinical Psychology* **59** 751–766.

LINDLEY, D. (1982). The role of randomization in inference. *PSA*: *Proceedings of the Biennial Meeting of the Philosophy of Science Association* **2** 431–446.

MAHALANOBIS, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences* (*Calcutta*) **2** 49–55.

MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. Academic Press, London. MR0560319

MORGAN, K. L. and RUBIN, D. B. (2012). Rerandomization to improve covariate balance in experiments. *Ann. Statist.* **40** 1263–1282. MR2985950

MORGAN, K. L. and RUBIN, D. B. (2015). Rerandomization to balance tiers of covariates. *J. Amer. Statist. Assoc.* **110** 1412–1421. MR3449036

MORRIS, C. (1979). A finite selection model for experimental design of the Health Insurance study. *J. Econometrics* **11** 43–61.

PAPINEAU, D. (1994). The virtues of randomization. *British J. Philos. Sci.* **45** 437–450, 712–715. MR1292321

RAVAUD, P., GIRAUDEAU, B., LOGEART, I., LARGUIER, J. S., ROLLAND, D., TREVES, R., EULLER-ZIEGLER, L., BANNWARTH, B. and DOUGADOS, M. (2004). Management of osteoarthritis (OA) with an unsupervised home based exercise programme and/or patient administered assessment tools. A cluster randomised controlled trial with a 2 × 2 factorial design. *Ann. Rheum. Dis.* **63** 703–708.

ROSENBAUM, P. R. and RUBIN, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Amer. Statist.* **39** 33–38.

ROSENBERGER, W. F. and SVERDLOV, O. (2008). Handling covariates in the design of clinical trials. *Statist. Sci.* **23** 404–419. MR2483911

RUBIN, D. B. (1976). Multivariate matching methods that are equal percent bias reducing. I. Some examples. *Biometrics* **32** 109–120. MR0400555

RUBIN, D. B. (2008). Comment: The design and analysis of gold standard randomized experiments [MR2655714]. *J. Amer. Statist. Assoc.* **103** 1350–1353. MR2655717

RUBIN, D. B. and THOMAS, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *J. Amer. Statist. Assoc.* **95** 573–585.

SEIDENFELD, T. (1982). *Levi on the Dogma of Randomization in Experiments* (H. E. Kyburg, JR. and I. Levi, eds.) 263–291. Springer, Berlin.

WORRALL, J. (2010). Evidence: Philosophy of science meets medicine. *J. Eval. Clin. Pract.* **16** 356–362.

WU, C. F. J. and HAMADA, M. S. (2009). *Experiments: Planning, Analysis, and Optimization*, 2nd ed. Wiley, Hoboken, NJ. MR2583259

XU, Z. and KALBFLEISCH, J. D. (2013). Repeated randomization and matching in multi-arm trials. *Biometrics* **69** 949–959. MR3146790

YATES, F. (1937). The design and analysis of factorial experiments. Imperial Bureau of Soil Sciences—Technical Communication. No. 35, Harpenden.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1977–1996
DOI: 10.1214/16-AOAS961
© Institute of Mathematical Statistics, 2016

PREDICTING MELBOURNE AMBULANCE DEMAND USING KERNEL WARPING[1]

BY ZHENGYI ZHOU AND DAVID S. MATTESON

Cornell University

Predicting ambulance demand accurately in fine resolutions in space and time is critical for ambulance fleet management and dynamic deployment. Typical challenges include data sparsity at high resolutions and the need to respect complex urban spatial domains. To provide spatial density predictions for ambulance demand in Melbourne, Australia, as it varies over hourly intervals, we propose a predictive spatio-temporal kernel warping method. To predict for each hour, we build a kernel density estimator on a sparse set of the most similar data from relevant past time periods (labeled data), but warp these kernels to a larger set of past data irregardless of time periods (point cloud). The point cloud represents the spatial structure and geographical characteristics of Melbourne, including complex boundaries, road networks and neighborhoods. Borrowing from manifold learning, kernel warping is performed through a graph Laplacian of the point cloud and can be interpreted as a regularization toward, and a prior imposed for, spatial features. Kernel bandwidth and degree of warping are efficiently estimated via cross-validation, and can be made time- and/or location-specific. Our proposed model gives significantly more accurate predictions compared to a current industry practice, an unwarped kernel density estimation and a time-varying Gaussian mixture model.

REFERENCES

AGGARWAL, C. C. (2003). A framework for diagonosing changes in evolving data streams. In *ACM SIGMOD International Conference on Management of Data* 575–586. ACM, New York.

BELKIN, M. and NIYOGI, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput*. **15** 1373–1396.

BELKIN, M. and NIYOGI, P. (2004). Semi-supervised learning on Riemannian manifolds. *Mach. Learn*. **56** 209–239.

BELKIN, M. and NIYOGI, P. (2005). Towards a theoretical foundation for Laplacian-based manifold methods. In *Learning Theory*. *Lecture Notes in Computer Science* **3559** 486–500. Springer, Berlin. MR2203282

BELKIN, M., NIYOGI, P. and SINDHWANI, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res*. **7** 2399–2434. MR2274444

BOUSQUET, O., CHAPELLE, O. and HEIN, M. (2005). Measure based regularization. In *Advances in Neural Information Processing Systems*. MIT Press, Cambridge.

CACOULLOS, T. (1966). Estimation of a multivariate density. *Ann. Inst. Statist. Math*. **18** 179–189. MR0210255

Key words and phrases. Emergency medical service, kernel density estimation, manifold learning, graph Laplacian.

CHANNOUF, N., L'ECUYER, P., INGOLFSSON, A. and AVRAMIDIS, A. N. (2007). The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta. *Health Care Manag. Sci.* **10** 25–45.

DIEBOLD, F. and MARIANO, R. (1995). Comparing predictive accuracy. *J. Bus. Econom. Statist.* **13** 253–263.

DIGGLE, P. J. (2003). *Statistical Analysis of Spatial Point Patterns*, 2nd ed. Arnold, London.

DONOHO, D. L. and GRIMES, C. (2005). Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data. In *Proceedings of the National Academy of Sciences* **102**. National Academy of Sciences, Washington DC.

DUONG, T. and HAZELTON, M. L. (2005). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scand. J. Stat.* **32** 485–506. MR2204631

ERTÖZ, L., STEINBACH, M. and KUMAR, V. (2003). Finding clusters of different sizes, shapes and densities in noisy, high dimensional data. In *Proceedings of the SIAM International Conference on Data Mining* 47–58. SIAM, Philadelphia.

ESTER, M., KRIEGEL, H. P., SANDER, J. and XU, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noice. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 226–231. ACM, New York.

FREY, B. J. and DUECK, D. (2007). Clustering by passing messages between data points. *Science* **315** 972–976. MR2292174

GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548

GOLDBERG, J. B. (2004). Operations research methods for the deployment of emergency service vehicles. *EMS Management Journal* **1** 20–39.

GOOD, I. J. (1952). Rational decisions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **14** 107–114. MR0077033

GOOGLE MAPS (2015). Map of Melbourne, Australia. Web.

GRAY, A. G. and MOORE, A. W. (2003). Nonparametric density estimation: Toward computational tractability. In *Proceedings of the SIAM International Conference on Data Mining*. SIAM, Philadelphia.

MATTESON, D. S., MCLEAN, M. W., WOODARD, D. B. and HENDERSON, S. G. (2011). Forecasting emergency medical service call arrival rates. *Ann. Appl. Stat.* **5** 1379–1406. MR2849778

MERRIS, R. (1994). Laplacian matrices of graphs: A survey. *Linear Algebra Appl.* **197/198** 143–176. MR1275613

MØLLER, J. and WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. *Monographs on Statistics and Applied Probability* **100**. Chapman & Hall/CRC, Boca Raton, FL. MR2004226

NAKAYA, T. and YANO, K. (2010). Visualising crime clusters in a space–time cube: An exploratory data analysis approach using space–time kernel density estimation and scan statistics. *Transactions in GIS* **14** 223–239.

NG, A., JORDAN, M. and WEISS, Y. (2001). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems* 849–856. MIT Press, Cambridge.

RAMSAY, T. (2002). Spline smoothing over difficult regions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 307–319. MR1904707

REGIS, R. G. and SHOEMAKER, C. A. (2007). A stochastic radial basis function method for the global optimization of expensive functions. *INFORMS J. Comput.* **19** 497–509. MR2364007

REGIS, R. G. and SHOEMAKER, C. A. (2009). Parallel stochastic global optimization using radial basis functions. *INFORMS J. Comput.* **21** 411–426. MR2546962

RIPLEY, B. D. and RASSON, J.-P. (1977). Finding the edge of a Poisson forest. *J. Appl. Probab.* **14** 483–491. MR0451339

ROWEIS, S. T. and SAUL, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* **290** 2323–2326.

SCOTT, D. W. (1992). *Multivariate Density Estimation*: *Theory*, *Practice*, *and Visualization*. Wiley, New York. MR1191168

SETZLER, H., SAYDAM, C. and PARK, S. (2009). EMS call volume predictions: A comparative study. *Comput*. *Oper*. *Res*. **36** 1843–1851.

SHI, J. and MALIK, J. (2000). Normalized cuts and image segmentation. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **22** 888–905.

SINDHWANI, V., NIYOGI, P. and BELKIN, M. (2005). Beyond the point cloud: From transductive to semi-supervised learning. In *Proceedings of the* 22*nd International Conference on Machine Learning* 824–831. ACM, New York.

SMOLA, A. J. and KONDOR, R. (2003). Kernels and regularization on graphs. In *Learning Theory and Kernel Machines*, *Lecture Notes in Computer Science* 144–158. Springer, Berlin.

TENEBAUM, J. B., DE SILVA, V. and LANGFORD, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290** 2319–2323.

VAN DER MAATEN, L. J. P., POSTMA, E. O. and VAN DEN HERIK, H. J. (2009). Dimensionality reduction: A comparative review. *J*. *Mach*. *Learn*. *Res*. **10** 66–71.

VILE, J. L., GILLARD, J. W., HARPER, P. R. and KNIGHT, V. A. (2012). Predicting ambulance demand using singular spectrum analysis. *Journal of the Operations Research Society* **63** 1556–1565.

WAND, M. P. and JONES, M. C. (1994). Multivariate plug-in bandwidth selection. *Comput*. *Statist*. **9** 97–116. MR1280754

WOOD, S. N., BRAVINGTON, M. V. and HEDLEY, S. L. (2008). Soap film smoothing. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **70** 931–955. MR2530324

WOODWORTH, J. T., MOHLER, G. O., BERTOZZI, A. L. and BRANTINGHAM, P. J. (2014). Non-local crime density estimation incorporating housing information. *Philos*. *Trans*. *R*. *Soc*. *Lond*. *Ser*. *A Math*. *Phys*. *Eng*. *Sci*. **372** 20130403, 15. MR3268065

ZHANG, Z., CHEN, D., LIU, W., RACINE, J. S., ONG, S. H., CHENG, Y., ZHAO, G. and JIANG, Q. (2011). Nonparametric evaluation of dynamic disease risk: A spatio-temporal kernel approach. *PLoS ONE* **6**.

ZHOU, Z. and MATTESON, D. S. (2015). Predicting ambulance demand: A spatio-temporal kernel approach. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York.

ZHOU, D., BOUSQUET, O., LAL, T. N., WESTON, J. and SCHOELKOPF, B. (2003). Learning with local and global consistency. In *Advances in Neural Information Processing Systems*. MIT Press, Cambridge.

ZHOU, Z., MATTESON, D. S., WOODARD, D. B., HENDERSON, S. G. and MICHEAS, A. C. (2015). A spatio-temporal point process model for ambulance demand. *J*. *Amer*. *Statist*. *Assoc*. **110** 6–15. MR3338482

ZHU, X., KANDOLA, J., GHAHRAMAMI, Z. and LAFFERTY, J. (2005). Nonparametric transforms of graph kernels for semi-supervised learning. In *Advances in Neural Information Processing Systems*. MIT Press, Cambridge.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1997–2020
DOI: 10.1214/16-AOAS962

MAXIMIZING THE INFORMATION CONTENT OF A BALANCED MATCHED SAMPLE IN A STUDY OF THE ECONOMIC PERFORMANCE OF GREEN BUILDINGS[1]

BY CINAR KILCIOGLU AND JOSÉ R. ZUBIZARRETA

Columbia University

Buildings have a major impact on the environment through excessive use of resources, such as energy and water, and large carbon dioxide emissions. In this paper we revisit a previously published study about the economics of environmentally sustainable buildings and estimate the effect of green building practices on market rents. For this, we use new matching methods that take advantage of the clustered structure of the buildings data. We propose a general framework for matching in observational studies and specific matching methods within this framework that simultaneously achieve three goals: (i) maximize the information content of a matched sample (and, in some cases, also minimize the variance of a difference-in-means effect estimator); (ii) form the matches using a flexible matching structure (such as a one-to-many/many-to-one structure); and (iii) directly attain covariate balance as specified—before matching—by the investigator. To our knowledge, existing matching methods are only able to achieve, at most, two of these goals simultaneously. Also, unlike most matching methods, the proposed methods do not require estimation of the propensity score or other dimensionality reduction techniques, although with the proposed methods these can be used as additional balancing covariates in the context of (iii). Using these matching methods, we find that green buildings have 3.3% higher rental rates per square foot than otherwise similar buildings without green ratings—a moderately larger effect than the one found by the prior study.

REFERENCES

ARONOW, P. M. and SAMII, C. (2016). Does regression produce representative estimates of causal effects? *Amer. J. Polit. Sci.* **60** 250–267.

BAIOCCHI, M. (2011). Designing robust studies using propensity score and prognostic score matching. Chapter 3 in Methodologies for Observational Studies of Health Care Policy, Dissertation, Department of Statistics, The Wharton School, Univ. Pennsylvania, Philadelphia, PA.

BAIOCCHI, M., SMALL, D. S., LORCH, S. and ROSENBAUM, P. R. (2010). Building a stronger instrument in an observational study of perinatal care for premature infants. *J. Amer. Statist. Assoc.* **105** 1285–1296. MR2796550

BERTSIMAS, D. (2014). Statistics and Machine Learning via a Modern Optimization Lens. The 2014–2015 Philip McCord Morse Lecture.

BIXBY, R. and ROTHBERG, E. (2007). Progress in computational mixed integer programming—A look back from the other side of the tipping point. *Ann. Oper. Res.* **149** 37–41. MR2313358

Key words and phrases. Causal inference, matched sampling, observational studies, propensity score.

CHAN, K. C. G., YAM, S. C. P. and ZHANG, Z. (2016). Globally efficient nonparametric inference of average treatment effects by empirical balancing calibration weighting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 673–700.

COCHRAN, W. G. (1965). The planning of observational studies of human populations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **128** 234–266.

COCHRAN, W. and RUBIN, D. (1973). Controlling bias in observational studies: A review. *Sankhya* **35** 417–446.

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

DIAMOND, A. and SEKHON, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Rev. Econ. Stat.* **95** 932–945.

EICHHOLTZ, P., KOK, N. and QUIGLEY, J. M. (2010). Doing well by doing good? Green office buildings. *Am. Econ. Rev.* **100** 2492–2509.

FOGARTY, C., MIKKELSEN, M., GAIESKI, D. and SMALL, D. (2016). Discrete optimization for interpretable study populations and randomization inference in an observational study of severe sepsis mortality. *J. Amer. Statist. Assoc.* **111** 447–458.

GRAHAM, B. S., DE XAVIER PINTO, C. C. and EGEL, D. (2012). Inverse probability tilting for moment condition model with missing data. *Rev. Econ. Stud.* **79** 1053–1079. MR2986390

HAINMUELLER, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Polit. Anal.* **20** 25–46.

HANSEN, B. B. (2004). Full matching in an observational study of coaching for the SAT. *J. Amer. Statist. Assoc.* **99** 609–618. MR2086387

HANSEN, B. B. (2007). Flexible, optimal matching for observational studies. *R News* **7** 18–24.

HANSEN, B. B. and BOWERS, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statist. Sci.* **23** 219–236. MR2516821

HANSEN, B. B. and KLOPFER, S. O. (2006). Optimal full matching and related designs via network flows. *J. Comput. Graph. Statist.* **15** 609–627. MR2280151

HANSEN, B. B., ROSENBAUM, P. R. and SMALL, D. S. (2014). Clustered treatment assignments and sensitivity to unmeasured biases in observational studies. *J. Amer. Statist. Assoc.* **109** 133–144. MR3180552

HARTMAN, E., GRIEVE, R., RAMSAHAI, R. and SEKHON, J. S. (2015). From sample average treatment effect to population average treatment effect on the treated: Combining experimental with observational studies to estimate population treatment effects. *J. Roy. Statist. Soc. Ser. A* **178** 757–778. MR3348358

HAVILAND, A., NAGIN, D. and ROSENBAUM, P. (2007). Combining propensity score matching and group-based trajectory analysis in an observational study. *Psychol. Methods* **12** 247.

HILL, J. (2008). Discussion of research using propensity-score matching: Comments on “A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003” by Peter Austin, *Statistics in Medicine* [MR2439882]. *Stat. Med.* **27** 2055–2061. MR2439884

HSU, J. Y., ZUBIZARRETA, J. R., SMALL, D. S. and ROSENBAUM, P. R. (2015). Strong control of the familywise error rate in observational studies that discover effect modification by exploratory methods. *Biometrika* **102** 767–782. MR3431552

IACUS, S. M., KING, G. K. and PORRO, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Polit. Anal.* **20** 1–24.

IMAI, K. and RATKOVIC, M. (2015). Robust estimation of inverse probability weights of marginal structural models. *J. Amer. Statist. Assoc.* **110** 1013–1023. MR3420680

IMBENS, G. W. (2015). Matching methods in practice: Three examples. *J. Hum. Resour.* **50** 373–419.

IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—For Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. MR3309951

KALTON, G. (1968). Standardization: A technique to control for extraneous variables. *Appl. Statist.* **17** 118–136. MR0234599

KEELE, L., TITIUNIK, R. and ZUBIZARRETA, J. R. (2015). Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. *J. Roy. Statist. Soc. Ser. A* **178** 223–239. MR3291769

KILCIOGLU, C. and ZUBIZARRETA, J. R. (2016). Supplement to "Maximizing the information content of a balanced matched sample in a study of the economic performance of green buildings." DOI:10.1214/16-AOAS962SUPP.

LEHMANN, E. L. (2006). *Nonparametrics*: *Statistical Methods Based on Ranks*, 1st ed. Springer, New York. MR2279708

LI, F., MORGAN, K. L. and ZASLAVSKY, A. M. (2016). Balancing covariates via propensity score weighting. Working paper.

LI, Y. P., PROPERT, K. J. and ROSENBAUM, P. R. (2001). Balanced risk set matching. *J. Amer. Statist. Assoc.* **96** 870–882. MR1946360

LINDEROTH, J. T. and LODI, A. (2010). MILP software. In *Wiley Encyclopedia of Operations Research and Management Science* (J. J. Cochran, L. A. Cox, P. Keskinocak and J. P. Kharoufeh, eds.). Wiley, New York.

LU, B. (2005). Propensity score matching with time-dependent covariates. *Biometrics* **61** 721–728. MR2196160

NEMHAUSER, G. L. (2013). Integer programming: Global impact. EURO INFORMS, July 2013.

NIKOLAEV, A. G., JACOBSON, S. H., CHO, W. K. T., SAUPPE, J. J. and SEWELL, E. C. (2013). Balance optimization subset selection (BOSS): An alternative approach for causal inference with observational data. *Oper. Res.* **61** 398–412. MR3046118

PIMENTEL, S. D., KELZ, R. R., SILBER, J. H. and ROSENBAUM, P. R. (2015). Large, sparse optimal matching with refined covariate balance in an observational study of the health outcomes produced by new surgeons. *J. Amer. Statist. Assoc.* **110** 515–527. MR3367244

ROSENBAUM, P. R. (1987). Model-based direct adjustment. *J. Amer. Statist. Assoc.* **82** 387–394.

ROSENBAUM, P. R. (1989). Optimal matching for observational studies. *J. Amer. Statist. Assoc.* **84** 1024–1032.

ROSENBAUM, P. (1991). Discussing hidden bias in observational studies. *Arch. Intern. Med.* **115** 901–905.

ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. Springer, New York. MR1899138

ROSENBAUM, P. R. (2005). Heterogeneity and causality: Unit heterogeneity and design sensitivity in observational studies. *Amer. Statist.* **59** 147–152. MR2133562

ROSENBAUM, P. R. (2010). *Design of Observational Studies*. Springer, New York. MR2561612

ROSENBAUM, P. R. (2014). Weighted M-statistics with superior design sensitivity in matched observational studies with multiple controls. *J. Amer. Statist. Assoc.* **109** 1145–1158. MR3265687

ROSENBAUM, P. R. (2015). How to see more in observational studies: Some new quasi-experimental devices. *Annual Review of Statistics and Its Application* **2** 21–48.

ROSENBAUM, P. R., ROSS, R. N. and SILBER, J. H. (2007). Minimum distance matched sampling with fine balance in an observational study of treatment for ovarian cancer. *J. Amer. Statist. Assoc.* **102** 75–83. MR2345534

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

ROSENBAUM, P. R. and RUBIN, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Amer. Statist.* **39** 33–38.

ROSENBAUM, P. R. and SILBER, J. (2001). Matching and thick description in an observational study of mortality after surgery. *Biostatistics* **2** 217–232.

ROSENBAUM, P. R. and SILBER, J. H. (2009). Amplification of sensitivity analysis in matched observational studies. *J. Amer. Statist. Assoc.* **104** 1398–1405. MR2750570

RUBIN, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *J. Amer. Statist. Assoc.* **74** 318–328.

RUBIN, D. B. (2008). For objective causal inference, design trumps analysis. *Ann. Appl. Stat.* **2** 808–840. MR2516795

SILBER, J. H., ROSENBAUM, P. R., KELZ, R. R., GASKIN, D. J., LUDWIG, J. M., ROSS, R. N., NIKNAM, B. A., HILL, A., WANG, M., EVEN-SHOSHAN, O. and FLEISHER, L. A. (2015). Examining causes of racial disparities in general surgical mortality: Hospital quality versus patient risk. *Med. Care* **53** 619–629.

STUART, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statist. Sci.* **25** 1–21. MR2741812

TRASKIN, M. and SMALL, D. (2011). Defining the study population for an observational study to ensure sufficient overlap: A tree approach. *Statistics in Biosciences* **3** 94–118.

TUKEY, J. W. (1986). Sunset salvo. *Amer. Statist.* **40** 72–76.

WESTON, S. and CALAWAY, R. (2014). Getting Started with `doParallel` and `foreach`.

YANG, D., SMALL, D. S., SILBER, J. H. and ROSENBAUM, P. R. (2012). Optimal matching with minimal deviation from fine balance in a study of obesity and surgical outcomes. *Biometrics* **68** 628–636. MR2959630

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

ZUBIZARRETA, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *J. Amer. Statist. Assoc.* **110** 910–922. MR3420672

ZUBIZARRETA, J. R. and KILCIOGLU, C. (2016). *designmatch*: Construction of Optimally Matched Samples for Randomized Experiments and Observational Studies that are Balanced by Design R package version 0.2.0.

ZUBIZARRETA, J. R., PAREDES, R. D. and ROSENBAUM, P. R. (2014). Matching for balance, pairing for heterogeneity in an observational study of the effectiveness of for-profit and not-for-profit high schools in Chile. *Ann. Appl. Stat.* **8** 204–231. MR3191988

ZUBIZARRETA, J. R., REINKE, C. E., KELZ, R. R., SILBER, J. H. and ROSENBAUM, P. R. (2011). Matching for several sparse nominal variables in a case–control study of readmission following surgery. *Amer. Statist.* **65** 229–238. MR2867507

ZUBIZARRETA, J. R., SMALL, D. S., GOYAL, N. K., LORCH, S. and ROSENBAUM, P. R. (2013). Stronger instruments via integer programming in an observational study of late preterm birth outcomes. *Ann. Appl. Stat.* **7** 25–50. MR3086409

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2021–2046
DOI: 10.1214/16-AOAS963

MODELING CONCURRENCY AND SELECTIVE MIXING IN HETEROSEXUAL PARTNERSHIP NETWORKS WITH APPLICATIONS TO SEXUALLY TRANSMITTED DISEASES[1]

BY RYAN ADMIRAAL AND MARK S. HANDCOCK

Murdoch University and University of California, Los Angeles

Network-based models for sexually transmitted disease transmission rely on initial partnership networks incorporating structures that may be related to risk of infection. In particular, initial networks should reflect the level of concurrency and attribute-based selective mixing observed in the population of interest. We consider momentary degree distributions as measures of concurrency and propensities for people of certain types to form partnerships with each other as a measure of attribute-based selective mixing. Estimation of momentary degree distributions and mixing patterns typically relies on cross-sectional survey data, and, in the context of heterosexual networks, we describe how this results in two sets of reports that need not be consistent with each other. The reported momentary degree distributions and mixing totals are related through a series of constraints, however. We provide a method to incorporate those in jointly estimating momentary degree distributions and mixing totals. We develop a method to simulate heterosexual networks consistent with these momentary degree distributions and mixing totals, applying it to data obtained from the National Longitudinal Study of Adolescent Health. We first use the momentary degree distributions and mixing totals as mean value parameters to estimate the natural parameters for an exponential-family random graph model and then use a Markov chain Monte Carlo algorithm to simulate person-level heterosexual partnership networks.

REFERENCES

ADIMORA, A. A. and SCHOENBACH, V. J. (2002). Contextual factors and the black–white disparity in heterosexual HIV transmission. *Epidemiology* **13** 707–712.

ADIMORA, A. A. and SCHOENBACH, V. J. (2005). Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. *J. Infect. Dis.* **191** S115–S122.

ADIMORA, A. A., SCHOENBACH, V. J. and DOHERTY, I. A. (2006). HIV and African Americans in the southern United States: Sexual networks and social context. *Sex. Transm. Dis.* **33** S39–S45.

ADIMORA, A. A., SCHOENBACH, V. J. and DOHERTY, I. A. (2007). Concurrent sexual partnerships among men in the United States. *Am. J. Publ. Health* **97** 2230–2237.

ADMIRAAL, R. and HANDCOCK, M. S. (2016). Supplement to "Modeling concurrency and selective mixing in heterosexual partnership networks with applications to sexually transmitted diseases." DOI:10.1214/16-AOAS963SUPP.

ANDERSON, R. M. (1992). Some aspects of sexual behavior and the potential demographic impact of AIDS in developing countries. *Social Science and Medicine* **34** 271–280.

Key words and phrases. Heterosexual partnership networks, exponential-family random graph models, constrained maximum likelihood estimation, National Longitudinal Survey of Adolescent Health.

Anderson, R. M., Gupta, S. and Ng, W. (1990). The significance of sexual partner contact networks for the transmission dynamics of HIV. *J. Acquir. Immune Defic. Syndr.* **3** 417–429.

Aral, S. O., Hughes, J. P., Stoner, B., Whittington, W., Handsfield, H. H., Anderson, R. M. and Holmes, K. K. (1999). Sexual mixing patterns in the spread of gonococcal and chlamydial infections. *Am. J. Publ. Health* **89** 825–833.

Barndorff-Nielsen, O. (2014). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester. MR3221776

Busenberg, S. and Castillo-Chavez, C. (1989). Interaction, pair formation and force of infection terms in sexually transmitted diseases. In *Mathematical and Statistical Approaches to AIDS Epidemiology*. *Lecture Notes in Biomathematics* **83** 289–300. Springer, Berlin. MR1040595

Carnegie, N. B. and Morris, M. (2011). Size matters: Concurrency and the epidemic potential of HIV in small networks. *PLoS One* **7** e43048.

Castillo-Chavez, C. and Blythe, S. P. (1989). Mixing framework for social/sexual behavior. In *Mathematical and Statistical Approaches to AIDS Epidemiology* (C. Castillo-Chavez, ed.). *Lecture Notes in Biomathematics* **83** 275–288. Springer, Berlin. MR1040594

Centers for Disease Control and Prevention (2015). Sexually Transmitted Diseases: Data & Statistics. Available at http://www.cdc.gov/std/stats/default.htm, 2015. Accessed: December 1, 2015.

Chick, S. E., Adams, A. L. and Koopman, J. S. (2000). Analysis and simulation of a stochastic, discrete-individual model of STD transmission with partnership concurrency. *Math. Biosci.* **166** 45–68.

Demographic and Health Surveys Program (2015). HIV/AIDS Survey Indicators Database. Available at http://hivdata.dhsprogram.com, 2015. Accessed: December 1, 2015.

Doherty, I. A., Shiboski, S., Ellen, J. M., Adimora, A. A. and Padian, N. S. (2006). Sexual bridging socially and over time: A simulation model exploring the relative effects of mixing and concurrency on viral sexually transmitted infection transmission. *Sex. Transm. Dis.* **33** 368–373.

Eaton, J. W., Hallett, T. B. and Garnett, G. P. (2011). Concurrent sexual partnerships and primary HIV infection: A critical interaction. *AIDS Behav.* **15** 687–692.

Eaton, J. W., McGrath, N. and Newell, M.-L. (2012). Unpacking the recommended indicator for concurrent sexual partnerships. *AIDS* **26** 1037–1039.

Garnett, G. and Anderson, R. M. (1993a). Contact tracing and the estimation of sexual mixing patterns: The epidemiology of Gonococcal infections. *Sex. Transm. Dis.* **20** 181–191.

Garnett, G. P. and Anderson, R. M. (1993b). Factors controlling the spread of HIV in heterosexual communities in developing countries: Patterns of mixing between different age and sexual activity classes. *Philosophical Transactions*: *Biological Sciences* **342** 137–159.

Garnett, G. P., Hughes, J. P., Anderson, R. M., Stoner, B. P., Aral, S. O., Whittington, W. L., Handsfield, H. H. and Holmes, K. K. (1996). Sexual mixing patterns of patients attending sexually transmitted diseases clinics. *Sex. Transm. Dis.* **23** 248–257.

Ghalanos, A. and Theussls, S. (2012). Rsolnp: General Non-linear Optimization Using Augmented Lagrange Multiplier Method. Available at CRAN.R-project.org/package=Rsolnp. Version 1.14.

Ghani, A. C. and Garnett, G. P. (2000). Risks of acquiring and transmitting sexually transmitted diseases in sexual partner networks. *Sex. Transm. Dis.* **27** 579–587.

Ghani, A. C., Swinton, J. and Garnett, G. P. (1997). The role of sexual partnership networks in the epidemiology of gonorrhea. *Sex. Transm. Dis.* **24** 45–56.

Glynn, J. R., Dube, A., Kayuni, N., Floyd, S., Molesworth, A., Parrott, F., French, N. and Crampin, A. C. (2012). Measuring concurrency: An empirical study of different methods in a large population-based survey in northern Malawi and evaluation of the UNAIDS guidelines. *AIDS* **26** 977–985.

GOODREAU, S. M. (2011). A decade of modelling research yields considerable evidence for the importance of concurrency: A response to Sawers and Stillwaggon. *Journal of the International AIDS Society* **14** 1–7.

GOODREAU, S. M., CASSELS, S., KASPRZYK, D., MONTAÑO, D. E., GREEK, A. and MORRIS, M. (2012). Concurrent partnerships, acute infection and HIV epidemic dynamics among young adults in Zimbabwe. *AIDS Behav.* **6** 312–322.

GRULICH, A. E. and ZABLOTSKA, I. (2010). Commentary: Probability of HIV transmission through anal intercourse. *Int. J. Epidemiol.* **39** 1064–1065.

GUPTA, S., ANDERSON, R. M. and MAY, R. M. (1989). Networks of sexual contacts: Implications for the pattern of spread of HIV. *AIDS* **3** 807–817.

HAMILTON, D. T., HANDCOCK, M. S. and MORRIS, M. (2008). Degree distributions in sexual networks: A framework for evaluating evidence. *Sex. Transm. Dis.* **35** 30–40.

HAMILTON, D. T. and MORRIS, M. (2015). The racial disparities in STI in the U.S.: Concurrency, STI prevalence, and heterogeneity in partner selection. *Epidemics* **11** 56–61.

HANDCOCK, M. S. (2003). Assessing degeneracy in statistical models of social networks. Working paper, Center for Statistics and the Social Sciences, Univ. of Washington.

HANDCOCK, M. S. and GILE, K. J. (2010). Modeling networks from sampled data. *Ann. Appl. Stat.* **40** 285–327.

HANDCOCK, M. S., RENDALL, M. S. and CHEADLE, J. E. (2005). Improved regression estimation of a multivariate relationship with population data on the bivariate relationship. *Sociol. Method.* **35** 291–334.

HANDCOCK, M. S., HUNTER, D. R., BUTTS, C. T., GOODREAU, S. M. and MORRIS, M. (2003). statnet: Software tools for the Statistical Modeling of Network Data. Seattle, WA, 2003. Available at http://statnetproject.org.

HANDCOCK, M. S., HUNTER, D. R., BUTTS, C. T., GOODREAU, S. M., KRIVITSKY, P. N. and MORRIS, M. (2013). ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks. The Statnet Project (http://www.statnet.org), 2013. Available at CRAN.R-project.org/package=ergm. R package version 3.1-0.

HARRIS, K. M., HALPERN, C. T., WHITSEL, E., HUSSEY, J., TABOR, J., ENTZEL, P. and UDRY, J. R. (2009). The National Longitudinal Study of Adolescent Health: Research Design [www document]. Technical report, Carolina Population Center, University of North Carolina at Chapel Hill, Available at: http://www.cpc.unc.edu/projects/addhealth/design.

HELLERINGER, S., MKANDAWIRE, J. and KOHLER, H.-P. (2014). A new approach to measuring partnership concurrency and its association with HIV risk in couples. *AIDS Behav.* **18** 2291–2301.

HUDSON, C. (1993). Concurrent partnerships could cause AIDS epidemics. *International Journal of STD and AIDS* **4** 349–353.

HUNTER, D. R. and HANDCOCK, M. S. (2006). Inference in curved exponential family models for networks. *J. Comput. Graph. Statist.* **15** 565–583. MR2291264

HUNTER, D. R., HANDCOCK, M. S., BUTTS, C. T., GOODREAU, S. M. and MORRIS, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *J. Stat. Softw.* **24** 1–11.

HYMAN, J. M. and STANLEY, E. A. (1988). Using mathematical models to understand the AIDS epidemic. *Math. Biosci.* **90** 415–473. MR0958152

JACQUEZ, J. A., SIMON, C. P. and KOOPMAN, J. (1989). Structured mixing: Heterogeneous mixing by the definition of activity groups. In *Mathematical and Statistical Approaches to AIDS Epidemiology* (C. Castillo-Chavez, ed.). *Lecture Notes in Biomathematics* **83** 301–315. Springer, Berlin. MR1040596

JOHNSON, L. F., DORRINGTON, R. E., BRADSHAW, D., PILLAY-VAN WYK, V. and REHLE, T. M. (2009). Sexual behaviour patterns in South Africa and their association with the spread of HIV: Insights from a mathematical model. *Demogr. Res. Monogr.* **21** 289–339.

JULIAN, D., BOUCHARD, C., GAGNON, M. and POMERLEAU, A. (1992). Insider's views of marital sex: A dyadic analysis. *J. Sex Res.* **29** 343–360.

KINSEY, A. C., POMEROY, W. B. and MARTIN, C. E. (1948). *Sexual Behavior in the Human Male*. W. B. Saunders Company, Philadelphia.

KOOPMAN, J. S., CHICK, S. E., RIOLO, C. S., ADAMS, A. L., WILSON, M. L. and BECKER, M. P. (2000). Modeling contact networks and infection transmission in geographic and social space using GERMS. *Sex. Transm. Dis.* **27** 617–626.

KRETZSCHMAR, M. and CARAËL, M. (2012). Is concurrency driving HIV transmission in Sub-Saharan African sexual networks? The significance of sexual partnership typology. *AIDS Behav.* **16** 1746–1752.

KRETZSCHMAR, M. and MORRIS, M. (1996). Measures of concurrency in networks and the spread of infectious disease. *Math. Biosci.* **133** 165–195.

KRIVITSKY, P. N. (2012). Exponential-family random graph models for valued networks. *Electron. J. Stat.* **6** 1100–1128. MR2988440

KRIVITSKY, P. N. and HANDCOCK, M. S. (2014). A separable model for dynamic networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 29–46. MR3153932

MAH, T. L. and HALPERIN, D. T. (2010). Concurrent sexual partnerships and the HIV epidemics in Africa: Evidence to move forward. *AIDS Behav.* **14** 11–16.

MORIN, B. R., PERRINGS, C., LEVIN, S. and KINZIG, A. (2014). Disease risk mitigation: The equivalence of two selective mixing strategies on aggregate contact patterns and resulting epidemic spread. *J. Theoret. Biol.* **363** 262–270. MR3278717

MORRIS, M. (1991). A log-linear modeling framework for selective mixing. *Math. Biosci.* **2** 349–377.

MORRIS, M. (1994). Epidemiology and social networks: Modeling structured diffusion. In *Advances in Social Network Analysis*: *Research in the Social and Behavioral Sciences* (S. Wasserman and J. Galaskiewicz, eds.) 26–52. Sage Publications, Thousand Oaks.

MORRIS, M. (1995). Data driven network models for the spread of infectious disease. In *Epidemic Models*: *Their Structure and Relation to Data* (D. Mollison, ed.) 302–322. Cambridge Univ. Press, Cambridge.

MORRIS, M. (1997). Sexual networks and HIV. *AIDS* **11** S209–S216.

MORRIS, M., EPSTEIN, H. and WAWER, M. (2010). Timing is everything: International variations in historical sexual partnership concurrency and HIV prevalence. *PLoS ONE* **5** 31–33.

MORRIS, M., GOODREAU, S. M. and MOODY, J. (2007). Sexual networks, concurrency, and STD/HIV. In *Sexually Transmitted Diseases*, 4th ed. (K. K. Holmes, P. F. Sparling, W. E. Stamm, P. Piot, J. N. Wasserheit, L. Corey and D. H. Watts, eds.) 109–125. McGraw-Hill, New York.

MORRIS, M. and KRETZSCHMAR, M. (1995). Concurrent partnerships and transmission dynamics in networks. *Social Networks* **17** 299–318.

MORRIS, M. and KRETZSCHMAR, M. (1997). Concurrent partnerships and the spread of HIV. *AIDS* **5** 641–648.

MORRIS, M. and KRETZSCHMAR, M. (2000). A microsimulation study of the effect of concurrent partnerships on the spread of HIV in Uganda. *Math. Popul. Stud.* **8** 109–133. The population dynamics of the HIV epidemic: projections. MR1806009

MORRIS, M., KURTH, A. E., HAMILTON, D. T., MOODY, J. and WAKEFIELD, S. (2009). Concurrent partnerships and HIV prevalence disparities by race: Linking science and public health practice. *Am. J. Publ. Health* **99** 1023–1031.

OCHS, E. P. and BINIK, Y. M. (1999). The use of couple data to determine the reliability of self-reported sexual behavior. *J. Sex Res.* **36** 374–384.

R CORE TEAM (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. Available at http://www.R-project.org/.

RENDALL, M. S., ADMIRAAL, R., DEROSE, A., DIGIULIO, P., HANDCOCK, M. S. and RACIOPPI, F. (2008). Population constraints on pooled surveys in demographic hazard modeling. *Stat. Methods Appl.* **17** 519–539. MR2447573

SAWERS, L. (2013). Measuring and modelling concurrency. *Journal of the International AIDS Society* **16** 1–20.

SEAL, D. W. (1997). Interpartner concordance of self-reported sexual behavior among college dating couples. *The Journal of Sex Research* **34** 39–55.

SHALIZI, C. R. and RINALDO, A. (2013). Consistency under sampling of exponential random graph models. *Ann. Statist.* **41** 508–535. MR3099112

SNIJDERS, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociol. Method.* **31** 361–95.

UNAIDS REFERENCE GROUP ON ESTIMATES, MODELLING, AND PROJECTIONS (2009). Consultation on Concurrent Sexual Partnerships: Recommendations from a meeting of the UNAIDS Reference Group on Estimates, Modelling and Projections held in Nairobi, Kenya, April 20–21st 2009.

WATTS, C. H. and MAY, R. M. (1992). The influence of concurrent partnerships on the dynamics of HIV/AIDS. *Math. Biosci.* **108** 89–104.

WORLD HEALTH ORGANIZATION (2013). Number of people (all ages) living with HIV: Estimates by WHO region. Available at http://apps.who.int/gho/data/view.main.22100WHO?, 2013. Accessed: December 1, 2015.

YE, Y. (1987). Interior algorithms for linear, quadratic, and linearly constrained non-linear programming Ph.D. Thesis, Stanford Univ., Dept. of EES.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2047–2074
DOI: 10.1214/16-AOAS964
In the Public Domain

INFERRING ROOTED POPULATION TREES USING ASYMMETRIC NEIGHBOR JOINING

BY YONGLIANG ZHAI AND ALEXANDRE BOUCHARD-CÔTÉ[1]

University of British Columbia

We introduce a new inference method to estimate evolutionary distances for any two populations to their most recent common ancestral population using single-nucleotide polymorphism allele frequencies. Our model takes fixation into consideration, making it nonreversible, and guarantees that the distribution of reconstructed ancestral frequencies is contained on the interval [0, 1]. To scale this method to large numbers of populations, we introduce the asymmetric neighbor joining algorithm, an efficient method for reconstructing rooted bifurcating nonclock trees. Asymmetric neighbor joining provides a scalable rooting method applicable to any nonreversible evolutionary modeling setups. We explore the statistical properties of asymmetric neighbor joining, and demonstrate its accuracy on synthetic data. We validate our method by reconstructing rooted phylogenetic trees from the Human Genome Diversity Panel data. Our results are obtained without using an outgroup, and are consistent with the prevalent recent single-origin model.

REFERENCES

BALDING, D. J. and NICHOLS, R. A. (1995). A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* **96** 3–12.

BATTISTUZZI, F. U., FILIPSKI, A., HEDGES, S. B. and KUMAR, S. (2010). Performance of relaxed-clock methods in estimating evolutionary divergence times and their credibility intervals. *Mol. Biol. Evol.* **27** 1289–1300.

BENNER, P., BAČÁK, M. and BOURGUIGNON, P.-Y. (2014). Point estimates in phylogenetic reconstructions. *Bioinformatics* **30** i534–i540.

BILLERA, L. J., HOLMES, S. P. and VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. MR1867931

BRYANT, D., BOUCKAERT, R., FELSENSTEIN, J., ROSENBERG, N. A. and ROYCHOUDHURY, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* **29** 1917–1932.

CANN, H. M., DE TOMA, C., CAZES, L., LEGRAND, M.-F., MOREL, V., PIOUFFRE, L., BODMER, J., BODMER, W. F., BONNE-TAMIR, B., CAMBON-THOMSEN, A. et al. (2002). A human genome diversity cell line panel. *Science* **296** 261–262.

CAVALLI-SFORZA, L. L. and FELDMAN, M. W. (2003). The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.* **33 Suppl** 266–275.

CHAKERIAN, J. and HOLMES, S. (2012). Computational tools for evaluating phylogenetic and hierarchical clustering trees. *J. Comput. Graph. Statist.* **21** 581–599. MR2970909

Key words and phrases. Asymmetric neighbor-joining algorithm, fixation and drift, phylogenetics, population histories, rooted tree inference, single-nucleotide polymorphism.

EDWARDS, A. and CAVALLI-SFORZA, L. (1964). Reconstruction of evolutionary trees. *Systematics Association Publ.* **6** 67–76.

EWENS, W. J. (1973). Conditional diffusion processes in population genetics. *Theor. Popul. Biol.* **4** 21–30.

FELSENSTEIN, J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* **25** 471–492.

FELSENSTEIN, J. (1981). Evolutionary trees from gene frequencies and quantitative characters: Finding maximum likelihood estimates. *Evolution* **35** 1229–1242.

FELSENSTEIN, J. (1983). Statistical inference of phylogenies. *J. R. Stat. Soc., A* **146** 246–272.

FELSENSTEIN, J. (1989). PHYLIP—Phylogeny inference package (Version 3.2). *Cladistics* **5** 164–166.

FELSENSTEIN, J. (2004). *Inferring Phytogenies*. Sinauer, Sunderland, Massachusetts.

GASCUEL, O. (1997). Concerning the NJ algorithm and its unweighted version, UNJ. In *Mathematical Hierarchies and Biology* (*Piscataway, NJ*, 1996). *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **37** 149–170. AMS, Providence, RI. MR1600536

GRAY, R. D. and ATKINSON, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* **426** 435–439.

GRAY, R. D., DRUMMOND, A. J. and GREENHILL, S. J. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323** 479–483.

HASEGAWA, M., KISHINO, H. and YANO, T.-A. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22** 160–174.

HERNANDEZ, R. D., KELLEY, J. L., ELYASHIV, E., MELTON, S. C., AUTON, A., MCVEAN, G., SELLA, G., PRZEWORSKI, M. et al. (2011). Classic selective sweeps were rare in recent human evolution. *Science* **331** 920–924.

HUELSENBECK, J. P., BOLLBACK, J. P. and LEVINE, A. M. (2002). Inferring the root of a phylogenetic tree. *Syst. Biol.* **51** 32–43.

HUELSENBECK, J. P., RONQUIST, F., NIELSEN, R. and BOLLBACK, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294** 2310–2314.

IWABE, N., KUMA, K-I., HASEGAWA, M., OSAWA, S. and MIYATA, T. (1989). Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. USA* **86** 9355–9359.

JENKINS, P. A. and SPANO, D. (2015). Exact simulation of the Wright–Fisher diffusion. preprint. Available at arXiv:1506.06998.

KUHNER, M. K. and FELSENSTEIN, J. (1994). A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* **11** 459–468.

LI, S., PEARL, D. K. and DOSS, H. (2000). Phylogenetic tree construction using Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **95** 493–508.

LI, J. Z., ABSHER, D. M., TANG, H., SOUTHWICK, A. M., CASTO, A. M., RAMACHANDRAN, S., CANN, H. M., BARSH, G. S., FELDMAN, M., CAVALLI-SFORZA, L. L. and MYERS, R. M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319** 1100–1104.

LIPO, C. P. (2006). *Mapping Our Ancestors*: *Phylogenetic Approaches in Anthropology and Prehistory*. Transaction Publishers. New Brunswick and London.

MAU, B., NEWTON, M. A. and LARGET, B. (1999). Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics* **55** 1–12. MR1705672

NEI, M. (1972). Genetic distance between populations. *Amer. Nat.* **106** 283–292.

NICHOLS, J. and WARNOW, T. (2008). Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* **2** 760–820.

NICHOLSON, G., SMITH, A. V., JÓNSSON, F., GÚSTAFSSON, O., STEFÁNSSON, K. and DONNELLY, P. (2002). Assessing population differentiation and isolation from single-nucleotide polymorphism data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 695–715. MR1979384

OUTLAW, D. C. and RICKLEFS, R. E. (2011). Rerooting the evolutionary tree of malaria parasites. *Proc. Natl. Acad. Sci. USA* **108** 13183–13187.

OWEN, M. and PROVAN, J. S. (2011). A fast algorithm for computing geodesic distances in tree space. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (*TCBB*) **8** 2–13.

PARADIS, E. (2012). *Analysis of Phylogenetics and Evolution with R*, 2nd ed. Springer, New York. MR2883250

PARADIS, E., CLAUDE, J. and STRIMMER, K. (2004). Ape: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20** 289–290.

PEARSON, T., HORNSTRA, H. M., SAHL, J. W., SCHAACK, S., SCHUPP, J. M., BECKSTROM-STERNBERG, S. M., O'NEILL, M. W., PRIESTLEY, R. A., CHAMPION, M. D., BECKSTROM-STERNBERG, J. S., KERSH, G. J., SAMUEL, J. E., MASSUNG, R. F. and KEIM, P. (2013). When outgroups fail; phylogenomics of rooting the emerging pathogen, Coxiella burnetii. *Syst. Biol.* **62** 752–762.

PENNY, D. and HENDY, M. (1985). The use of tree comparison metrics. *Syst. Zool.* **34** 75–82.

PICKRELL, J. K. and PRITCHARD, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8** e1002967.

PICKRELL, J. K., PATTERSON, N., BARBIERI, C., BERTHOLD, F., GERLACH, L., GÜLDEMANN, T., KURE, B., MPOLOKA, S. W., NAKAGAWA, H., NAUMANN, C. et al. (2012). The genetic prehistory of southern Africa. *Nature Communications* **3** 1–6.

REVELL, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3** 217–223.

ROCH, S. (2006). A short proof that phylogenetic tree reconstruction by maximum likelihood is hard. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (*TCBB*) **3** 92.

ROYCHOUDHURY, A., FELSENSTEIN, J. and THOMPSON, E. A. (2008). A two-stage pruning algorithm for likelihood computation for a population tree. *Genetics* **180** 1095–1105.

SAITOU, N. and NEI, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4** 406–425.

SEMPLE, C. and STEEL, M. (2003). *Phylogenetics*. *Oxford Lecture Series in Mathematics and Its Applications* **24**. Oxford Univ. Press, Oxford. MR2060009

SIRÉN, J., HANAGE, W. P. and CORANDER, J. (2013). Inference on population histories by approximating infinite alleles diffusion. *Mol. Biol. Evol.* **30** 457–468.

SIRÉN, J., MARTTINEN, P. and CORANDER, J. (2011). Reconstructing population histories from single nucleotide polymorphism data. *Mol. Biol. Evol.* **28** 673–683.

SMEULDERS, M. J., BARENDS, T. R. M., POL, A., SCHERER, A., ZANDVOORT, M. H., UDVARHELYI, A., KHADEM, A. F., MENZEL, A., HERMANS, J., SHOEMAN, R. L. et al. (2011). Evolution of a new enzyme for carbon disulphide conversion by an acidothermophilic archaeon. *Nature* **478** 412–416.

SONG, Y. S. and STEINRÜCKEN, M. (2012). A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. *Genetics* **190** 1117–1129.

SUKUMARAN, J. and HOLDER, M. T. (2010). DendroPy: A Python library for phylogenetic computing. *Bioinformatics* **26** 1569–1571.

SWOFFORD, D. L., OLSEN, G. J., WADDELL, P. J. and HILLIS, D. M. (1996). Phylogenetic inference. In *Molecular Systematics* (M. D. Hillis and C. Moritz, eds.) 407–514. Sinauer Associates, Sunderland.

TAVARÉ, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. In *Some Mathematical Questions in Biology—DNA Sequence Analysis* (*New York*, 1984). *Lectures Math. Life Sci.* **17** 57–86. AMS, Providence, RI. MR0846877

WANG, L., BOUCHARD-CÔTÉ, A. and DOUCET, A. (2015). Bayesian phylogenetic inference using a combinatorial sequential Monte Carlo method. *J. Amer. Statist. Assoc.* **110** 1362–1374. MR3449032

WEIR, B. S. and COCKERHAM, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38** 1358–1370.

WHEELER, W. C. (1990). Nucleic acid sequence phylogeny and random outgroups. *Cladistics* **6** 363–367.

YANG, Z., GOLDMAN, N. and FRIDAY, A. (1995). Maximum likelihood trees from DNA sequences: A peculiar statistical estimation problem. *Systematic Biology* **44** 384–399.

YANG, Z. and RANNALA, B. (1997). Bayesian phylogenetic inference using DNA sequences: A Markov chain Monte Carlo method. *Mol. Biol. Evol.* **14** 717–724.

ZHAI, Y. and BOUCHARD-CÔTÉ, A. (2016). Supplement to "Inferring rooted population trees using asymmetric neighbor joining." DOI:10.1214/16-AOAS964SUPP.

ZHARKIKH, A. and LI, W. H. (1995). Estimation of confidence in phylogeny: The complete-and-partial bootstrap technique. *Mol. Phylogenet. Evol.* **4** 44–63.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2075–2101
DOI: 10.1214/16-AOAS965
© Institute of Mathematical Statistics, 2016

MODELLING THE EFFECT OF THE EL NIÑO-SOUTHERN OSCILLATION ON EXTREME SPATIAL TEMPERATURE EVENTS OVER AUSTRALIA

BY HUGO C. WINTER[1], JONATHAN A. TAWN AND SIMON J. BROWN[2]

EDF Energy R&D UK Centre, Lancaster University and Met Office Hadley Centre

When assessing the risk posed by high temperatures, it is necessary to consider not only the temperature at separate sites but also how many sites are expected to be hot at the same time. Hot events that cover a large area have the potential to put a great strain on health services and cause devastation to agriculture, leading to high death tolls and much economic damage. South-eastern Australia experienced a severe heatwave in early 2009; 374 people died in the state of Victoria and Melbourne recorded its highest temperature since records began in 1859 [Nairn and Fawcett (2013)]. One area of particular interest in climate science is the effect of large-scale climatic phenomena, such as the El Niño-Southern Oscillation (ENSO), on extreme temperatures. Here, we develop a framework based upon extreme value theory to estimate the effect of ENSO on extreme temperatures across Australia. This approach permits us to estimate the change in temperatures with ENSO at important sites, such as Melbourne, and also whether we are more likely to observe hot temperatures over a larger spatial extent during a particular phase of ENSO. To this end, we design a set of measures that can be used to effectively summarise many important spatial aspects of an extreme temperature event. These measures are estimated using our extreme value framework and we validate whether we can accurately replicate the 2009 Australian heatwave, before using the model to estimate the probability of having a more severe event than has been observed.

REFERENCES

ALEXANDER, L. V. and ARBLASTER, J. M. (2009). Assessing trends in observed and modelled climate extremes over Australia in relation to future projections. *International Journal of Climatology* **29** 417–435.

AVILA, F. B., DONG, S., MENANG, K. P., RAJCZAK, J., RENOM, M., DONAT, M. G. and ALEXANDER, L. V. (2015). Systematic investigation of gridding-related scaling effects on annual statistics of daily temperature and precipitation maxima: A case study for south-east Australia. *Weather and Climate Extremes* **9** 6–16.

CAESAR, J., ALEXANDER, L. and VOSE, R. (2006). Large-scale changes in observe daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *Journal of Geophysical Research*: *Atmospheres* **111** 1–10.

CHAVEZ-DEMOULIN, V. and DAVISON, A. C. (2005). Generalized additive modelling of sample extremes. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **54** 207–222. MR2134607

Key words and phrases. Conditional extremes, covariates, El Niño-Southern Oscillation, extremal dependence, extreme temperature, severity-area-frequency curves, spatial extremes.

COLES, S. G. (1993). Regional modelling of extreme storms via max-stable processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **55** 797–816. MR1229882

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London. MR1932132

COLES, S. G., HEFFERNAN, J. E. and TAWN, J. A. (1999). Dependence measures for extreme value analyses. *Extremes* **2** 339–365.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York. MR1239641

DAVIS, R. A., KLÜPPELBERG, C. and STEINKOHL, C. (2013). Statistical inference for max-stable processes in space and time. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 791–819. MR3124792

DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27** 161–186. MR2963980

DAVISON, A. C. and SMITH, R. L. (1990). Models for exceedances over high thresholds. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **52** 393–442. MR1086795

DOMBRY, C., ÉYI-MINKO, F. and RIBATET, M. (2013). Conditional simulation of max-stable processes. *Biometrika* **100** 111–124. MR3034327

EASTOE, E. F. and TAWN, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **58** 25–45. MR2662232

HEFFERNAN, J. E. and RESNICK, S. I. (2007). Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.* **17** 537–571. MR2308335

HEFFERNAN, J. E. and TAWN, J. A. (2004). A conditional approach for multivariate extreme values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 497–546. MR2088289

HENRIQUES, A. G. and SANTOS, M. J. J. (1999). Regional drought distribution model. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere* **24** 19–22.

HUSER, R. and DAVISON, A. C. (2014). Space–time modelling of extreme events. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 439–461. MR3164873

JONES, D. A. and TREWIN, B. C. (2000). On the relationships between the El Nino-Southern Oscillation and Australian land surface temperature. *International Journal of Climatology* **20** 697–719.

KEEF, C., PAPASTATHOPOULOS, I. and TAWN, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *J. Multivariate Anal.* **115** 396–404. MR3004566

KENYON, J. and HEGERL, G. C. (2008). Influence of modes of climate variability on global temperature extremes. *Journal of Climate* **21** 3872–3889.

LEDFORD, A. W. and TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83** 169–187. MR1399163

MIN, S., CAI, W. and WHETTON, P. (2013). Influence of climate variability on seasonal extremes over Australia. *Journal of Geophysical Research: Atmospheres* **118** 643–654.

NAIRN, J. and FAWCETT, R. (2013). Defining heatwaves: Heatwave defined as a heat-impact event servicing all community and business sectors in Australia. Centre for Australian Weather and Climate Research, Technical Report, 060 1–96.

NORTHROP, P. J. and JONATHAN, P. (2011). Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics* **22** 799–809. MR2861046

PERKINS, S. E. and ALEXANDER, L. V. (2013). On the measurement of heat waves. *Journal of Climate* **26** 4500–4517.

SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5** 33–44. MR1947786

SELF, S. G. and LIANG, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* **82** 605–610. MR0898365

SMITH, R. L. (1990). Max-stable processes and spatial extremes. 1–32. Unpublished manuscript.

WANG, C. and PICAUT, J. (2004). Understanding ENSO physics—A review. *Geophysical Monograph Series* **147** 21–48.

WINTER, H. C. (2016). Extreme value modelling of heatwaves. Ph.D. thesis. Lancaster Univ.

WINTER, H. C. and TAWN, J. A. (2016). Modelling heatwaves in central France: A case study in extremal dependence. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **65** 345–365.

WINTER, H. C., TAWN, J. A. and BROWN, S. J. (2016). Detecting changing behaviour of heatwaves with climate change. Preprint.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2102–2129
DOI: 10.1214/16-AOAS966

THE SCREENING AND RANKING ALGORITHM FOR CHANGE-POINTS DETECTION IN MULTIPLE SAMPLES[1]

BY CHI SONG[2], XIAOYI MIN[2] AND HEPING ZHANG

Ohio State University, Georgia State University and Yale University

The chromosome copy number variation (CNV) is the deviation of genomic regions from their normal copy number states, which may associate with many human diseases. Current genetic studies usually collect hundreds to thousands of samples to study the association between CNV and diseases. CNVs can be called by detecting the change-points in mean for sequences of array-based intensity measurements. Although multiple samples are of interest, the majority of the available CNV calling methods are single sample based. Only a few multiple sample methods have been proposed using scan statistics that are computationally intensive and designed toward either common or rare change-points detection. In this paper, we propose a novel multiple sample method by adaptively combining the scan statistic of the screening and ranking algorithm (SaRa), which is computationally efficient and is able to detect both common and rare change-points. We prove that asymptotically this method can find the true change-points with almost certainty and show in theory that multiple sample methods are superior to single sample methods when shared change-points are of interest. Additionally, we report extensive simulation studies to examine the performance of our proposed method. Finally, using our proposed method as well as two competing approaches, we attempt to detect CNVs in the data from the Primary Open-Angle Glaucoma Genes and Environment study, and conclude that our method is faster and requires less information while our ability to detect the CNVs is comparable or better.

REFERENCES

ALKAN, C., COE, B. P. and EICHLER, E. E. (2011). Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **12** 363–376.

CAI, T. T., JENG, X. J. and JIN, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 629–662. MR2867452

CARTER, N. P. (2007). Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat. Genet.* **39** S16–S21.

DISKIN, S. J., LI, M., HOU, C., YANG, S., GLESSNER, J., HAKONARSON, H., BUCAN, M., MARIS, J. M. and WANG, K. (2008). Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* **36** e126.

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

FAN, Z., DROR, R. O., MILDORF, T. J., PIANA, S. and SHAW, D. E. (2015). Identifying localized changes in large systems: Change-point detection for biomolecular simulations. *Proc. Natl. Acad. Sci. USA* **112** 1–6.

Key words and phrases. Change-point detection, multi-sample inference, adaptive Fisher's method.

FISHER, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh, Chambers.

GONZALEZ, E., KULKARNI, H., BOLIVAR, H., MANGANO, A., SANCHEZ, R., CATANO, G., NIBBS, R. J., FREEDMAN, B. I., QUINONES, M. P., BAMSHAD, M. J. et al. (2005). The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307** 1434–1440.

HAO, N., NIU, Y. S. and ZHANG, H. (2013). Multiple change-point detection via a screening and ranking algorithm. *Statist. Sinica* **23** 1553–1572. MR3222810

HUANG, T., WU, B., LIZARDI, P. and ZHAO, H. (2005). Detection of DNA copy number alterations using penalized least squares regression. *Bioinformatics* **21** 3811–3817.

JENG, X. J., CAI, T. T. and LI, H. (2013). Simultaneous discovery of rare and common segment variants. *Biometrika* **100** 157–172. MR3034330

KORN, J. M., KURUVILLA, F. G., MCCARROLL, S. A., WYSOKER, A., NEMESH, J., CAWLEY, S., HUBBELL, E., VEITCH, J., COLLINS, P. J., DARVISHI, K. et al. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40** 1253–1260.

LENGAUER, C., KINZLER, K. W. and VOGELSTEIN, B. (1998). Genetic instabilities in human cancers. *Nature* **396** 643–649.

LI, J. and TSENG, G. C. (2011). An adaptively weighted statistic for detecting differential gene expression when combining multiple transcriptomic studies. *Ann. Appl. Stat.* **5** 994–1019. MR2840184

LITTELL, R. C. and FOLKS, J. L. (1971). Asymptotic optimality of Fisher's method of combining independent tests. *J. Amer. Statist. Assoc.* **66** 802–806. MR0312634

LITTELL, R. C. and FOLKS, J. L. (1973). Asymptotic optimality of Fisher's method of combining independent tests. II. *J. Amer. Statist. Assoc.* **68** 193–194. MR0375577

MCCARROLL, S. A. and ALTSHULER, D. M. (2007). Copy-number variation and association studies of human disease. *Nat. Genet.* **39** S37–S42.

MCCARROLL, S. A., HUETT, A., KUBALLA, P., CHILEWSKI, S. D., LANDRY, A., GOYETTE, P., ZODY, M. C., HALL, J. L., BRANT, S. R., CHO, J. H. et al. (2008). Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat. Genet.* **40** 1107–1112.

MERMEL, C. H., SCHUMACHER, S. E., HILL, B., MEYERSON, M. L., BEROUKHIM, R., GETZ, G. et al. (2011). GISTIC2. 0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12** R41.

NIU, Y. S. and ZHANG, H. (2012). The screening and ranking algorithm to detect DNA copy number variations. *Ann. Appl. Stat.* **6** 1306–1326. MR3012531

OLSHEN, A. B., VENKATRAMAN, E. S., LUCITO, R. and WIGLER, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5** 557–572.

POLLACK, J. R., SØRLIE, T., PEROU, C. M., REES, C. A., JEFFREY, S. S., LONNING, P. E., TIBSHIRANI, R., BOTSTEIN, D., BØRRESEN-DALE, A.-L. and BROWN, P. O. (2002). Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl. Acad. Sci. USA* **99** 12963–12968.

SEBAT, J., LAKSHMI, B., MALHOTRA, D., TROGE, J., LESE-MARTIN, C., WALSH, T., YAMROM, B., YOON, S., KRASNITZ, A., KENDALL, J. et al. (2007). Strong association of de novo copy number mutations with autism. *Science* **316** 445–449.

SIEGMUND, D., YAKIR, B. and ZHANG, N. R. (2011). Detecting simultaneous variant intervals in aligned sequences. *Ann. Appl. Stat.* **5** 645–668. MR2840169

SONG, C., MIN X. and ZHANG, H. (2016). Supplement to "The screening and ranking algorithm for change-points detection in multiple samples." DOI:10.1214/16-AOAS966SUPP.

STOUFFER, S. A., SUCHMAN, E. A., DEVINNEY, L. C., STAR, S. A. and WILLIAMS JR, R. M. (1949). *The American Soldier*: *Adjustment During Army Life*. Princeton Univ. Press, Princeton.

TIBSHIRANI, R. and WANG, P. (2008). Spatial smoothing and hot spot detection for CGH data using the fused lasso. *Biostatistics* **9** 18–29.

VENKATRAMAN, E. S. and OLSHEN, A. B. (2007). A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23** 657–663.

VERT, J. and BLEAKLEY, K. (2010). Fast detection of multiple change-points shared by many signals using group LARS. In *Advances in Neural Information Processing Systems* 23 (J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel and A. Culotta, eds.) 2343–2351. Curran Associates, Red Hook.

WANG, K., LI, M., HADLEY, D., LIU, R., GLESSNER, J., GRANT, S. F. A., HAKONARSON, H. and BUCAN, M. (2007). PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research* **17** 1665–1674.

XIAO, F., MIN, X. and ZHANG, H. (2015). Modified screening and ranking algorithm for copy number variation detection. *Bioinformatics* **31** 1341–1348.

YAO, Y.-C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statist. Probab. Lett.* **6** 181–189. MR0919373

YAO, Y.-C. and AU, S. T. (1989). Least-squares estimation of a step function. *Sankhyā Ser. A* **51** 370–381. MR1175613

YU, K., LI, Q., BERGEN, A. W., PFEIFFER, R. M., ROSENBERG, P. S., CAPORASO, N., KRAFT, P. and CHATTERJEE, N. (2009). Pathway analysis by adaptive combination of P-values. *Genet. Epidemiol.* **33** 700–709.

ZHANG, S., CHEN, H.-S. and PFEIFFER, R. M. (2013). A combined p-value test for multiple hypothesis testing. *J. Statist. Plann. Inference* **143** 764–770. MR3003888

ZHANG, N. R., SIEGMUND, D. O., JI, H. and LI, J. Z. (2010). Detecting simultaneous changepoints in multiple sequences. *Biometrika* **97** 631–645. MR2672488

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2130–2156
DOI: 10.1214/16-AOAS967

COX REGRESSION WITH EXCLUSION FREQUENCY-BASED WEIGHTS TO IDENTIFY NEUROIMAGING MARKERS RELEVANT TO HUNTINGTON'S DISEASE ONSET

BY TANYA P. GARCIA[1] AND SAMUEL MÜLLER[2]

Texas A&M University and University of Sydney

Biomedical studies of neuroimaging and genomics collect large amounts of data on a small subset of subjects so as to not miss informative predictors. An important goal is identifying those predictors that provide better visualization of the data and that could serve as cost-effective measures for future clinical trials. Identifying such predictors is challenging, however, when the predictors are naturally interrelated and the response is a failure time prone to censoring. We propose to handle these challenges with a novel variable selection technique. Our approach casts the problem into several smaller dimensional settings and extracts from this intermediary step the relative importance of each predictor through data-driven weights called exclusion frequencies. The exclusion frequencies are used as weights in a weighted Lasso, and results yield low false discovery rates and a high geometric mean of sensitivity and specificity. We illustrate the method's advantages over existing ones in an extensive simulation study, and use the method to identify relevant neuroimaging markers associated with Huntington's disease onset.

REFERENCES

AYLWARD, E. H. (2007). Change in MRI striatal volumes as a biomarker in preclinical Huntington's disease. *Brain Res. Bull.* **72** 152–158.

AYLWARD, E. H., NOPOULOS, P. C., ROSS, C. A., LANGBEHN, D., PIERSON, R. K., MILLS, J. A., JOHNSON, H., MAGNOTTA, V., JUHL, A., PAULSEN, J. S. and THE PREDICT-HD INVESTIGATORS AND COORDINATORS OF THE HUNTINGTON STUDY GROUP (2011). Longitudinal change in regional brain volumes in prodromal Huntington disease. *J. Neurol. Neurosurg. Psychiatry* **82** 405–410.

AYLWARD, E. H., LIU, D., NOPOULOS, P. C., ROSS, C. A., PIERSON, R. K., MILLS, J. A., LONG, J. D., PAULSEN, J. S. and THE PREDICT-HD INVESTIGATORS, AND COORDINATORS OF THE HUNTINGTON STUDY GROUP (2012). Striatal volume contributes to the prediction of onset of Huntington disease in incident cases. *Biological Psychiatry* **71** 822–828. PMID: 21907324, PMCID, PMC3237730.

BACH, F. (2008). Bolasso: Model consistent Lasso estimation through the bootstrap. In *Proceedings of the* 25*th International Conference on Machine Learning*, Helsinki, Finland. 2008.

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

BERGERSEN, L. C., GLAD, I. K. and LYNG, H. (2011). Weighted lasso with data integration. *Stat. Appl. Genet. Mol. Biol.* **10** Art. 39, 31. MR2837183

Key words and phrases. Exclusion frequency, model selection, neuroimaging, proportional hazards model, weighted lasso.

BUCKLAND, S. T., BURNHAM, K. P. and AUGUSTIN, N. H. (1997). Model selection: An integral part of inference. *Biometrics* **53** 603–619.

CHEN, C. H. and GEORGE, S. L. (1985). The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Stat. Med.* **4** 39–46.

COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758

HUNTINGTON'S DISEASE COLLABORATIVE RESEARCH GROUP (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72** 971–983.

FAN, J. and LI, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *Ann. Statist.* **30** 74–99. MR1892656

FARAGGI, D. and SIMON, R. (1998). Bayesian variable selection method for censored survival data. *Biometrics* **54** 1475–1485. MR1671590

GARCIA, T. P. and MÜLLER, S. (2014). Influence of measures of significance based weights in the weighted lasso. *J. Indian Soc. Agricultural Statist.* **68** 131–144. MR3242570

GARCIA, T. P., MÜLLER, S., CARROLL, R. J., DUNN, T. N., THOMAS, A. P., ADAMS, S. H., PILLAI, S. D. and WALZEM, R. L. (2013). Structured variable selection with q-values. *Biostatistics* **14** 695–707.

GARCIA, T. P., MÜLLER, S., CARROLL, R. J. and WALZEM, R. L. (2014). Identification of important regressor groups, subgroups and individuals via regularization methods: Application to gut microbiome data. *Bioinformatics* **30** 831–837.

GEORGIOU-KARISTIANIS, N., SCAHILL, R., TABRIZI, S. J., SQUITIERI, F. and AYLWARD, E. (2013). Structural MRI in Huntington's disease and recommendations for its potential use in clinical trials. *Neurosci. Biobehav. Rev.* **37** 480–490.

GONG, G. D. (1982). Cross-validation, the jacknife, and the bootstrap: Excess error estimation in forward logistic regression. Technical Report 192, Dept. of Statistics, Stanford Univ., 1–82.

GONG, G. (1986). Cross-validation, the jacknife, and the bootstrap: Excess error estimation in forward logistic regression. *J. Amer. Statist. Assoc.* **81** 108–113.

HICKS, S., ROSAS, H. D., BERNA, C., SCAHILL, R., DURMAS, E., ROOS, R. A. et al. (2010). PAW36 oculomotor deficits in presymptomatic and early Huntington's disease and their structural brain correlates. *J. Neurol. Neurosurg. Psychiatry* **81** e33.

HOBBS, N. Z., BARNES, J., FROST, C., HENLEY, S. M. D., WILD, E. J., MACDONALD, K., BARKER, R. A., SCAHILL, R. I., FOX, N. C. and TABRIZI, S. J. (2010). Onset and progression of pathologic atrophy in Huntington disease: A longitudinal MR imaging study. *Am. J. Neuroradiol.* **31** 1036–1041.

IBRAHIM, J. G., CHEN, M.-H. and MACEACHERN, S. N. (1999). Bayesian variable selection for proportional hazards models. *Canad. J. Statist.* **27** 701–717. MR1767142

JURGENS, C. K., VAN DE WIEL, L., VAN ES, A. C. G. M., GRIMBERGEN, Y. M., WITJES-ANE, M. N. W., VAN DER GROND, J. et al. (2008). Basal ganglia volume and clinical correlates in 'pre-clinical' Huntington's disease. *J. Neurol.* **255** 1785–1791.

KUBAT, M., HOLTE, R. C. and MATWIN, S. (1998). Machine learning for the detection of oil spills in satellite radar images. *Mach. Learn.* **30** 195–215.

LANGBEHN, D. R., BRINKMAN, R. R., FALUSH, D., PAULSEN, J. S., HAYDEN, M. R. and INTERNATIONAL HUNTINGTON'S DISEASE COLLABORATIVE GROUP (2004). A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. *Clin. Genet.* **65** 267–277.

LIN, W. and LV, J. (2013). High-dimensional sparse additive hazards regression. *J. Amer. Statist. Assoc.* **108** 247–264. MR3174617

MEINSHAUSEN, N. and BÜHLMANN, P. (2010). Stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 417–473. MR2758523

MÜLLER, S. and WELSH, A. H. (2005). Outlier robust model selection in linear regression. *J. Amer. Statist. Assoc.* **100** 1297–1310. MR2236443

MÜLLER, S. and WELSH, A. H. (2009). Robust model selection in generalized linear models. *Statist. Sinica* **19** 1155–1170. MR2536149

MÜLLER, S. and WELSH, A. H. (2010). On model selection curves. *Int. Stat. Rev.* **78** 240–256.

PAULSEN, J. S., LANGBEHN, D. R., STOUT, J. C., AYLWARD, E., ROSS, C. A., NANCE, M., GUTTMAN, M., JOHNSON, S., MCDONALD, M., BEGLINGER, L. J., DUFF, K., KAYSON, E., BIGLAN, K., SHOULSON, I., OAKES, D., HAYDEN, M. and COORDINATORS OF THE HUNTINGTON STUDY GROUP (2008). Detection of Huntington's disease decades before diagnosis: The Predict HD study. *J. Neurol. Neurosurg. Psychiatry* **79** 874–880.

PAULSEN, J. S., NOPOULOS, P. C., AYLWARD, E., ROSS, C. A., JOHNSON, H., MAGNOTTA, V. A., JUHL, A., PIERSON, R. K., MILLS, J., LANGBEHN, D. and NANCE, M. (2010). Striatal and white matter predictors of estimated diagnosis for Huntington disease. *Brain Res. Bull.* **82** 201–207.

ROSS, C. A. and TABRIZI, S. J. (2010). Huntington's disease: From molecular pathogenesis to clinical treatment. *Lancet Neurol.* **10** 83–98.

ROSS, C. A., PANTELYAT, A., KOGAN, J. and BRANDT, J. (2014). Determinants of functional disability in Huntington's disease: Role of cognitive and motor dysfunction. *Mov. Disord.* **29** 1351–1358.

SAUERBREI, W. and SCHUMACHER, M. (1992). A bootstrap resampling procedure for model building: Application to the Cox regression model. *Stat. Med.* **11** 2093–2109.

SHAH, R. D. and SAMWORTH, R. J. (2013). Variable selection with error control: Another look at stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 55–80. MR3008271

SIMON, N., FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* **39** 1–13.

SIMON, N., FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2013). A sparse-group lasso. *J. Comput. Graph. Statist.* **22** 231–245. MR3173712

STOREY, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Statist.* **31** 2013–2035. MR2036398

STOREY, J. D. and TIBSHIRANI, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100** 9440–9445. MR1994856

TABRIZI, S. J., REILMANN, R., ROOS, R. A. C., DURR, A., LEAVITT, B., OWEN, G., JONES, R., JOHNSON, H., CRAUFURD, D., HICKS, S. L., KENNARD, C., LANDWEHRMEYER, B., STOUT, J. C., BOROWSKY, B., SCAHILL, R. I., FROST, C., LANGBEHN, D. R. and TRACK-HD INVESTIGATORS (2012). Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: Analysis of 24 month observational data. *Lancet Neurol.* **11** 42–53.

TABRIZI, S. J., SCAHILL, R. I., OWEN, G., DURR, A., LEAVITT, B. R., ROOS, R. A., BOROWSKY, B., LANDWEHRMEYER, B., FROST, C., JOHNSON, H., CRAUFURD, D., REILMANN, R., STOUT, J. C., LANGBEHN, D. R. and TRACK-HD INVESTIGATORS (2013). Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: Analysis of 36-month observational data. *Lancet Neurol.* **12** 637–649.

TIBSHIRANI, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* **16** 385–395.

WANG, S., NAN, B., ROSSET, S. and ZHU, J. (2011). Random Lasso. *Ann. Appl. Stat.* **5** 468–485. MR2810406

WASSEF, S. N., WEMMIE, J., JOHNSON, C. P., JOHNSON, H., PAULSEN, J. S., LONG, J. D. and MAGNOTTA, V. A. (2015). T1ρ imaging in premanifest Huntington disease reveals changes associated with disease progression. *Mov. Disord.* **30** 1107–1114.

WITTEN, D. M. and TIBSHIRANI, R. (2010). Survival analysis with high-dimensional covariates. *Stat. Methods Med. Res.* **19** 29–51. MR2744491

YOUNES, L., RATNANATHER, J. T., BROWN, T., AYLWARD, E., NOPOULOS, P., JOHNSON, H., MAGNOTTA, V. A., PAULSEN, J. S., MARGOLIS, R. L., ALBIN, R. L., MILLER, M. I. and ROSS, C. A. (2014). Regionally selective atrophy of subcortical structures in prodromal HD as revealed by statistical shape analysis. *Hum. Brain Mapp.* **35** 792–809.

YU, B. (2013). Stability. *Bernoulli* **19** 1484–1500. MR3102560

ZHANG, H. H. and LU, W. (2007). Adaptive Lasso for Cox's proportional hazards model. *Biometrika* **94** 691–703. MR2410017

ZHANG, Y., LONG, J. D., MILLS, J. A., WARNER, J. H., LU, W., PAULSEN, J. S. and THE PREDICT-HD INVESTIGATORS OF THE HUNTINGTON STUDY GROUP, C. (2011). Indexing disease progression at study entry with individuals at-risk for Huntington disease. *Am. J. Med. Genet., Part B Neuropsychiatr. Genet.* **156B** 751–763.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2157–2181
DOI: 10.1214/16-AOAS968
In the Public Domain

BAYESIAN NONPARAMETRIC MULTIRESOLUTION ESTIMATION FOR THE AMERICAN COMMUNITY SURVEY

BY TERRANCE D. SAVITSKY

U.S. Bureau of Labor Statistics

Bayesian hierarchical methods implemented for small area estimation focus on reducing the noise variation in published government official statistics by borrowing information among dependent response values. Even the most flexible models confine parameters defined at the finest scale to link to each data observation in a one-to-one construction. We propose a Bayesian multiresolution formulation that utilizes an ensemble of observations at a variety of coarse scales in space and time to additively nest parameters we define at a finer scale, which serve as our focus for estimation. Our construction is motivated by and applied to the estimation of 1-year period employment totals, indexed by county, from statistics published at coarser areal domains and multi-year periods in the American Community Survey (ACS). We construct a nonparametric mixture of Gaussian processes as the prior on a set of regression coefficients of county-indexed latent functions over multiple survey years. We evaluate a modified Dirichlet process prior that incorporates county-year predictors as the mixing measure. Each county-year parameter of a latent function is estimated from multiple coarse-scale observations in space and time to which it links. The multiresolution formulation is evaluated on synthetic data and applied to the ACS.

REFERENCES

BRADLEY, J. R., WIKLE, C. K. and HOLAN, S. H. (2014). Bayesian spatial change of support for count-valued survey data. Available at http://adsabs.harvard.edu/abs/2014arXiv1405.7227B.

BRADLEY, J. R., WIKLE, C. K. and HOLAN, S. H. (2015). Spatio-temporal change of support with application to American Community Survey multi-year period estimates. *Stat* **4** 255–270. MR3414659

CELEUX, G., FORBES, F., ROBERT, C. P. and TITTERINGTON, D. M. (2006). Rejoinder to "Deviance information criteria for missing data models." *Bayesian Anal.* **1** 701–706 (electronic). MR2282197

DAWID, A. P. (1981). Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika* **68** 265–274. MR0614963

ESCOBAR, M. D. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. MR1340510

GELFAND, A. E. and DEY, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *J. Roy. Statist. Soc. Ser. B* **56** 501–514. MR1278223

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2015). *Bayesian Data Analysis*, 3rd ed. Chapman & Hall/CRC, Boca Raton, FL.

Key words and phrases. Survey sampling, small area estimation, latent models, Gaussian process, Dirichlet process, Bayesian hierarchical models, Markov chain Monte Carlo.

GELMAN, A. and RUBIN, D. (1992). Inference from iterative simulation using multiple sequences. *Statist*. *Sci*. **7** 457–511. Available at http://www.stat.columbia.edu/~gelman/research/published/itsim.pdf.

GHOSH, M., NATARAJAN, K., STROUD, T. W. F. and CARLIN, B. P. (1998). Generalized linear models for small-area estimation. *J*. *Amer*. *Statist*. *Assoc*. **93** 273–282. MR1614644

HAWALA, S. and LAHIRI, P. (2012). Hierarchical Bayes estimation of poverty rates. Technical report, U.S. Census Bureau—Small Area Income and Poverty Estimates. Available at https://www.census.gov/did/www/saipe/publications/files/hawalalahirishpl2012.pdf.

JONES, G. L., HARAN, M., CAFFO, B. S. and NEATH, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J*. *Amer*. *Statist*. *Assoc*. **101** 1537–1547. MR2279478

MÜLLER, P., QUINTANA, F. and ROSNER, G. L. (2011). A product partition model with regression on covariates. *J*. *Comput*. *Graph*. *Statist*. **20** 260–278. MR2816548

RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA. MR2514435

RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields*: *Theory and Applications*. *Monographs on Statistics and Applied Probability* **104**. Chapman & Hall/CRC, Boca Raton, FL. MR2130347

SÄRNDAL, C., SWENSSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. Springer, New York. MR1140409

SAVITSKY, T. D. (2016). Supplement to "Bayesian nonparametric multiresolution estimation for the American Community Survey." DOI:10.1214/16-AOAS968SUPP.

SAVITSKY, T. D. and MCCAFFREY, D. F. (2013). Bayesisan hierarchical multivariate formulation with factor analysis for nested ordinal data. *Psychometrika* **79** 275–302.

SAVITSKY, T. D. and PADDOCK, S. M. (2013). Bayesian nonparametric hierarchical modeling for multiple membership data in grouped attendance interventions. *Ann*. *Appl*. *Stat*. **7** 1074–1094. MR3113501

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

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

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2182–2202
DOI: 10.1214/16-AOAS970
© Institute of Mathematical Statistics, 2016

DYNAMIC SOCIAL NETWORKS BASED ON MOVEMENT[1]

BY HENRY R. SCHARF[*], MEVIN B. HOOTEN[†,*], BAILEY K. FOSDICK[*], DEVIN S. JOHNSON[‡], JOSH M. LONDON[‡] AND JOHN W. DURBAN[§]

Colorado State University[*], U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit[†], NOAA Alaska Fisheries Science Center[‡] and NOAA Southwest Fisheries Science Center[§]

Network modeling techniques provide a means for quantifying social structure in populations of individuals. Data used to define social connectivity are often expensive to collect and based on case-specific, *ad hoc* criteria. Moreover, in applications involving animal social networks, collection of these data is often opportunistic and can be invasive. Frequently, the social network of interest for a given population is closely related to the way individuals move. Thus, telemetry data, which are minimally invasive and relatively inexpensive to collect, present an alternative source of information. We develop a framework for using telemetry data to infer social relationships among animals. To achieve this, we propose a Bayesian hierarchical model with an underlying dynamic social network controlling movement of individuals via two mechanisms: an attractive effect and an aligning effect. We demonstrate the model and its ability to accurately identify complex social behavior in simulation, and apply our model to telemetry data arising from killer whales. Using auxiliary information about the study population, we investigate model validity and find the inferred dynamic social network is consistent with killer whale ecology and expert knowledge.

REFERENCES

ANDREWS, R. D., PITMAN, R. L. and BALLANCE, L. T. (2008). Satellite tracking reveals distinct movement patterns for type B and type C killer whales in the southern Ross sea, Antarctica. *Polar Biol*. **31** 1461–1468.

BAIRD, R. W. and WHITEHEAD, H. (2000). Social organization of mammal-eating killer whales: Group stability and dispersal patterns. *Can. J. Zool*. **78** 2096–2105.

BERLINER, L. M. (1996). Hierarchical Bayesian time series models. In *Maximum Entropy and Bayesian Methods* (*Santa Fe*, *NM*, 1995) (K. M. Hanson and R. N. Silver, eds.). *Fund. Theories Phys*. **79** 15–22. Kluwer Academic, Dordrecht. MR1446713

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236. MR0373208

BESAG, J. and KOOPERBERG, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* **82** 733–746. MR1380811

BRILLINGER, D. R. and STEWART, B. S. (1998). Elephant-seal movements: Modelling migration. *Canad. J. Statist*. **26** 431–443.

CODLING, E. A. and BODE, N. W. (2014). Copycat dynamics in leaderless animal group navigation. *Movement Ecology* **2** 11.

Key words and phrases. Dynamic social network, animal movement, *Orcinus orca*, hidden Markov model, Gaussian Markov random field.

CROFT, D. P., JAMES, R. and KRAUSE, J. (2008). *Exploring Animal Social Networks*. Princeton Univ. Press, Princeton, NJ.

DURANTE, D. and DUNSON, D. B. (2014). Nonparametric Bayes dynamic modelling of relational data. *Biometrika* **101** 883–898. MR3286923

DURBAN, J. W. and PITMAN, R. L. (2012). Antarctic killer whales make rapid, round-trip movements to subtropical waters: Evidence for physiological maintenance migrations? *Biol. Lett.* **8** 274–277.

DURBAN, J. W., FEARNBACH, H., BURROWS, D. G., YLITALO, G. M. and PITMAN, R. L. (2016). morphological and ecological evidence for two sympatric forms of Type B killer whale around the Antarctic peninsula. *Polar Biology* **April** 1–6.

FORESTER, J. D., IVES, A. R., TURNER, M. G., ANDERSON, D. P., FORTIN, D., BEYER, H. L., SMITH, D. W. and BOYCE, M. S. (2007). State-space models link elk movement patterns to landscape characteristics in Yellowstone National Park. *Ecol. Mono.* **77** 285–299.

FRANZ, M., MCLEAN, E., TUNG, J., ALTMANN, J. and ALBERTS, S. C. (2015). Self-organizing dominance hierarchies in a wild primate population. *Proceedings of the Royal Society B: Biological Sciences* **282** 20151512.

FURRER, R., GENTON, M. G. and NYCHKA, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.* **15** 502–523. MR2291261

GELFAND, A. E., DIGGLE, P., GUTTORP, P. and FUENTES, M. (2010). *Handbook of Spatial Statistics*, CRC Press, Boca Raton, FL.

GOLDENBERG, S. Z., DE SILVA, S., RASMUSSEN, H. B., DOUGLAS-HAMILTON, I. and WITTEMYER, G. (2014). Controlling for behavioural state reveals social dynamics among male African elephants, *Loxodonta africana*. *Anim. Behav.* **95** 111–119.

HANKS, E. M., HOOTEN, M. B., JOHNSON, D. S. and STERLING, J. T. (2011). Velocity-based movement modeling for individual and population level inference. *PLoS ONE* **6** e22795.

HANKS, E. M., SCHLIEP, E. M., HOOTEN, M. B. and HOETING, J. A. (2015). Restricted spatial regression in practice: Geostatistical models, confounding, and robustness under model misspecification. *Environmetrics* **26** 243–254. MR3340961

HOOTEN, M. B., JOHNSON, D. S., HANKS, E. M. and LOWRY, J. H. (2010). Agent-based inference for animal movement and selection. *J. Agric. Biol. Environ. Stat.* **15** 523–538. MR2788638

JOHNSON, D. S., LONDON, J. M. and KUHN, C. E. (2011). Bayesian inference for animal space use and other movement metrics. *J. Agric. Biol. Environ. Stat.* **16** 357–370. MR2843131

JOHNSON, D. S., LONDON, J. M., LEA, M.-A. and DURBAN, J. W. (2008). Continuous-time correlated random walk model for animal telemetry data. *Ecology* **89** 1208–1215.

JONSEN, I. D., FLEMMING, J. M. and MYERS, R. A. (2005). Robust state-space modeling of animal movement data. *Ecology* **86** 2874–2880.

KRAUSE, J., CROFT, D. P. and JAMES, R. (2007). Social network theory in the behavioural sciences: Potential applications. *Behav. Ecol. Sociobiol.* **62** 15–27.

LANGROCK, R., KING, R., MATTHIOPOULOS, J., THOMAS, L., FORTIN, D. and MORALES, J. M. (2012). Flexible and practical modeling of animal telemetry data: Hidden Markov models and extensions. *Ecology* **93** 2336–2342.

LANGROCK, R., HOPCRAFT, J. G. C., BLACKWELL, P. G., GOODALL, V., KING, R., NIU, M., PATTERSON, T. A., PEDERSEN, M. W., SKARIN, A. and SCHICK, R. S. (2014). Modelling group dynamic animal movement. *Methods in Ecology and Evolution* **5** 190–199.

LEMASSON, B. H., ANDERSON, J. J. and GOODWIN, R. A. (2013). Motion-guided attention promotes adaptive communications during social navigation. *Proceedings of the Royal Society B: Biological Sciences* **280** 20122003.

LEVIN, I. I., ZONANA, D. M., BURT, J. M. and SAFRAN, R. J. (2015). Performance of encounternet tags: Field tests of miniaturized proximity loggers for use on small birds. *PLoS ONE* **10** e0137242.

McClintock, B. T., Johnson, D. S., Hooten, M. B., Hoef, J. M. V. and Morales, J. M. (2014). When to be discrete: The importance of time formulation in understanding animal movement. *Movement Ecology* **2** 21.

Miller, P. J. O. (2006). Diversity in sound pressure levels and estimated active space of resident killer whale vocalizations. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* **192** 449–459.

Morales, J. M., Moorcroft, P. R., Matthiopoulos, J., Frair, J. L., Kie, J. G., Powell, R. A., Merrill, E. H. and Haydon, D. T. (2010). Building the bridge between animal movement and population dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365** 2289–2301.

Morin, P. A., Parsons, K. M., Archer, F. I., Ávila-Arcos, M. C., Barrett-Lennard, L. G., Dalla Rosa, L., Duchêne, S., Durban, J. W., Ellis, G. M., Ferguson, S. H., Ford, J. K., Ford, M. J., Garilao, C., Gilbert, M. T. P., Kaschner, K., Matkin, C. O., Petersen, S. D., Robertson, K. M., Visser, I. N., Wade, P. R., Ho, S. Y. W. and Foote, A. D. (2015). geographical and temporal dynamics of a global radiation and diversification in the killer whale. *Mol. Ecol.* **24** 3964–3979.

Parsons, K. M., Balcomb, K. C., Ford, J. K. B. and Durban, J. W. (2009). The social dynamics of southern resident killer whales and conservation implications for this endangered population. *Anim. Behav.* **77** 963–971.

Pinter-Wollman, N., Hobson, E. A., Smith, J. E., Edelman, A. J., Shizuka, D., de Silva, S., Waters, J. S., Prager, S. D., Sasaki, T., Wittemyer, G., Fewell, J. and McDonald, D. B. (2013). The dynamics of animal social networks: Analytical, conceptual, and theoretical advances. *Behavioral Ecology* art047.

Pitman, R. L. and Durban, J. W. (2010). Killer whale predation on penguins in Antarctica. *Polar Biol.* **33** 1589–1594.

Pitman, R. L. and Durban, J. W. (2012). Cooperative hunting behavior, prey selectivity and prey handling by pack ice killer whales (*Orcinus orca*), type B, in Antarctic Peninsula waters. *Mar. Mamm. Sci.* **28** 16–36.

Pitman, R. L. and Ensor, P. (2003). Three forms of killer whales (*Orcinus orca*) in Antarctic waters. *J. Cetacean Res. Manag.* **5** 131–140.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications. Monographs on Statistics and Applied Probability* **104**. Chapman & Hall, Boca Raton, FL. MR2130347

Russell, J. C., Hanks, E. M. and Haran, M. (2015). Dynamic models of animal movement with spatial point process interactions. *J. Agric. Biol. Environ. Stat.* 1–19.

Sarkar, P. and Moore, A. W. (2005). Dynamic social network analysis using latent space models. *ACM SIGKDD Explor. Newsl.* **7** 31–40.

Scharf, H. R., Hooten, M. B., Fosdick, B. K., Johnson, D. S., London, J. M. and Durban, J. W. (2016a). Supplement to "Dynamic social networks based on movement." DOI:10.1214/16-AOAS970SUPPA.

Scharf, H. R., Hooten, M. B., Fosdick, B. K., Johnson, D. S., London, J. M. and Durban, J. W. (2016b). Supplement to "Dynamic social networks based on movement." DOI:10.1214/16-AOAS970SUPPB.

Scharf, H. R., Hooten, M. B., Fosdick, B. K., Johnson, D. S., London, J. M. and Durban, J. W. (2016c). Supplement to "Dynamic social networks based on movement." DOI:10.1214/16-AOAS970SUPPC.

Sewell, D. K. and Chen, Y. (2015). Latent space models for dynamic networks. *J. Amer. Statist. Assoc.* **110** 1646–1657. MR3449061

Sih, A., Hanser, S. F. and McHugh, K. A. (2009). Social network theory: New insights and issues for behavioral ecologists. *Behav. Ecol. Sociobiol.* **63** 975–988.

Wey, T., Blumstein, D. T., Shen, W. and Jordán, F. (2008). Social network analysis of animal behaviour: A promising tool for the study of sociality. *Anim. Behav.* **75** 333–344.

WILLIAMS, R. and LUSSEAU, D. (2006). A killer whale social network is vulnerable to targeted removals. *Biol. Lett.* **2** 497–500.

WILLIAMS, R., TRITES, A. W. and BAIN, D. E. (2002). Behavioural responses of killer whales (*Orcinus orca*) to whale-watching boats: Opportunistic observations and experimental approaches. *J. Zool.* **256** 255–270.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2203–2232
DOI: 10.1214/16-AOAS971
© Institute of Mathematical Statistics, 2016

LOCALLY ADAPTIVE DYNAMIC NETWORKS[1]

BY DANIELE DURANTE AND DAVID B. DUNSON

University of Padova and Duke University

Our focus is on realistically modeling and forecasting dynamic networks of face-to-face contacts among individuals. Important aspects of such data that lead to problems with current methods include the tendency of the contacts to move between periods of slow and rapid changes, and the dynamic heterogeneity in the actors' connectivity behaviors. Motivated by this application, we develop a novel method for Locally Adaptive DYnamic (LADY) network inference. The proposed model relies on a dynamic latent space representation in which each actor's position evolves in time via stochastic differential equations. Using a state-space representation for these stochastic processes and Pólya-gamma data augmentation, we develop an efficient MCMC algorithm for posterior inference along with tractable procedures for online updating and forecasting of future networks. We evaluate performance in simulation studies, and consider an application to face-to-face contacts among individuals in a primary school.

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.

BARRAT, A. and CATTUTO, C. (2013). Temporal networks of face-to-face human interactions. In *Understanding Complex Systems* 191–216. Springer Science Business Media.

BUTTS, C. T. (2008). A relational event framework for social action. *Sociol. Method.* **38** 155–200.

CATTUTO, C., VAN DEN BROECK, W., BARRAT, A., COLIZZA, V., PINTON, J.-F. and VESPIGNANI, A. (2010). Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS ONE* **5** e11596.

CHOI, H. M. and HOBERT, J. P. (2013). The Polya-gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic. *Electron. J. Stat.* **7** 2054–2064. MR3091616

DESMARAIS, B. A. and CRANMER, S. J. (2012). Statistical mechanics of networks: Estimation and uncertainty. *Phys. A* **391** 1865–1876.

DUBOIS, C., BUTTS, C. T., MCFARLAND, D. and SMYTH, P. (2013). Hierarchical models for relational event sequences. *J. Math. Psych.* **57** 297–309. MR3137883

DUNSON, D. B. and XING, C. (2009). Nonparametric Bayes modeling of multivariate categorical data. *J. Amer. Statist. Assoc.* **104** 1042–1051. MR2562004

DURANTE, D. and DUNSON, D. B. (2014). Nonparametric Bayes dynamic modelling of relational data. *Biometrika* **101** 883–898. MR3286923

DURANTE, D., SCARPA, B. and DUNSON, D. B. (2014). Locally adaptive factor processes for multivariate time series. *J. Mach. Learn. Res.* **15** 1493–1522. MR3214789

DURBIN, J. and KOOPMAN, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* **89** 603–616. MR1929166

Key words and phrases. Face-to-face dynamic contact network, latent space, nested Gaussian process, online updating, Pólya-gamma, state-space model.

DURBIN, J. and KOOPMAN, S. J. (2012). *Time Series Analysis by State Space Methods*, 2nd ed. *Oxford Statistical Science Series* **38**. Oxford Univ. Press, Oxford. MR3014996

FIENBERG, S. E. and WASSERMAN, S. (1981). Categorical data analysis of single sociometric relations. *Sociol. Method.* **12** 156–192.

FOULDS, J., DUBOIS, C., ASUNCION, A. U., BUTTS, C. T. and SMYTH, P. (2011). A dynamic relational infinite feature model for longitudinal social networks. *Journal of Machine Learning Research Workshops & Proceedings* **15** 287–295.

FOURNET, J. and BARRAT, A. (2014). Contact patterns among high school students. *PLoS ONE* **9** e107878.

FRIEDMAN, J. H. (1991). Multivariate adaptive regression splines. *Ann. Statist.* **19** 1–141. MR1091842

FRUCHTERMAN, T. M. and REINGOLD, E. M. (1991). Graph drawing by force-directed placement. *Softw. Pract. Exp.* **21** 1129–1164.

GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–511.

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2013). *Bayesian Data Analysis*, 3rd ed. Taylor & Francis.

GEMMETTO, V., BARRAT, A. and CATTUTO, C. (2014). Mitigation of infectious disease at school: Targeted class closure vs school closure. *BMC Infect. Dis.* **14** 695.

GEORGE, E. I. and MCCULLOCH, R. E. (1993). Variable selection via Gibbs sampling. *J. Amer. Statist. Assoc.* **88** 881–889.

HANNEKE, S., FU, W. and XING, E. P. (2010). Discrete temporal models of social networks. *Electron. J. Stat.* **4** 585–605. MR2660534

HOFF, P. D. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems* 20 (J. C. Platt, D. Koller, Y. Singer and S. T. Roweis, eds.) 657–664. MIT Press, Cambridge.

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

HOLLAND, P. W. and LEINHARDT, S. (1977). A dynamic model for social networks. *J. Math. Sociol.* **5** 5–20. MR0446596

HUNTER, D. R., KRIVITSKY, P. N. and SCHWEINBERGER, M. (2012). Computational statistical methods for social network models. *J. Comput. Graph. Statist.* **21** 856–882. MR3005801

HUNTER, D. R., HANDCOCK, M. S., BUTTS, C. T., GOODREAU, S. M. and MORRIS, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *J. Stat. Softw.* **24** nihpa54860.

ISELLA, L., STEHLÉ, J., BARRAT, A., CATTUTO, C., PINTON, J.-F. and VAN DEN BROECK, W. (2011). What's in a crowd? Analysis of face-to-face behavioral networks. *J. Theoret. Biol.* **271** 166–180. MR2974883

KRIVITSKY, P. N. and HANDCOCK, M. S. (2014). A separable model for dynamic networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 29–46. MR3153932

MASTRANDREA, R., FOURNET, J. and BARRAT, A. (2015). Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE* **10** e0136497.

NEWMAN, M. E. J. (2003). Mixing patterns in networks. *Phys. Rev. E* (3) **67** 026126, 13. MR1975193

NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255

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

ROBINS, G. and PATTISON, P. (2001). Random graph models for temporal processes in social networks. *J. Math. Sociol.* **25** 5–41.

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.

SARKAR, P. and MOORE, A. W. (2005). Dynamic social network analysis using latent space models. *SIGKDD Explorations Newsletter* **7** 31–40.

SEWELL, D. K. and CHEN, Y. (2015). Latent space models for dynamic networks. *J. Amer. Statist. Assoc.* **110** 1646–1657. MR3449061

SNIJDERS, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociol. Method.* **31** 361–395.

SNIJDERS, T. A. B. (2005). Models for longitudinal network data. In *Models and Methods in Social Network Analysis* 215–247. Cambridge Univ. Press, Cambridge.

SNIJDERS, T. A. B., VAN DE BUNT, G. G. and STEGLICH, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Soc. Netw.* **32** 44–60.

STEHLÉ, J., VOIRIN, N., BARRAT, A., CATTUTO, C., ISELLA, L., PINTON, J.-F., QUAGGIOTTO, M., VAN DEN BROECK, W., RÉGIS, C., LINA, B. and VANHEMS, P. (2011). High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE* **6** e23176.

STEHLÉ, J., CHARBONNIER, F., PICARD, T., CATTUTO, C. and BARRAT, A. (2013). Gender homophily from spatial behavior in a primary school: A sociometric study. *Soc. Netw.* **35** 604–613.

VANHEMS, P., BARRAT, A., CATTUTO, C., PINTON, J.-F., KHANAFER, N., RÉGIS, C., KIM, B., COMTE, B. and VOIRIN, N. (2013). Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS ONE* **8** e73970.

WYATT, D., CHOUDHURY, T. and BILMES, J. A. (2008). Learning hidden curved exponential family models to infer face-to-face interaction networks from situated speech data. In *AAAI* 732–738.

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

XU, K. S. (2015). Stochastic block transition models for dynamic networks. *Journal of Machine Learning Research Workshops & Proceedings* **38** 1079–1087.

XU, K. S. and HERO, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing* **8** 552–562.

YANG, T., CHI, Y., ZHU, S., GONG, Y. and JIN, R. (2009). A Bayesian approach toward finding communities and their evolutions in dynamic social networks. In *Proceedings of the* 2009 *SIAM International Conference on Data Mining* 990–1001. Society for Industrial & Applied Mathematics (SIAM), Philadelphia.

YANG, T., CHI, Y., ZHU, S., GONG, Y. and JIN, R. (2011). Detecting communities and their evolutions in dynamic social networks—a Bayesian approach. *Mach. Learn.* **82** 157–189. MR3108191

ZHU, B. and DUNSON, D. B. (2013). Locally adaptive Bayes nonparametric regression via nested Gaussian processes. *J. Amer. Statist. Assoc.* **108** 1445–1456. MR3174720

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2233–2253
DOI: 10.1214/16-AOAS972

ESTIMATING ODDS RATIOS UNDER A CASE-BACKGROUND DESIGN WITH AN APPLICATION TO A STUDY OF SORAFENIB ACCESSIBILITY

BY JOHN H. SPIVACK AND BIN CHENG

Icahn School of Medicine at Mount Sinai and Columbia University

In certain epidemiologic studies such as those involving stress disorders, sexual harassment, alcohol addiction or epidemiological criminology, exposure data are readily available from cases but not from controls because it is socially inconvenient or even unethical to determine who qualifies as a true control subject. Consequently, it is impractical or even infeasible to use a case-control design to establish the case-exposure association in such situations. To address this issue, we propose a case-background design where in addition to a sample of exposure information from cases, an independent sample of exposure information from the background population is taken, without knowing the case status of the sampled subjects. We develop a semiparametric method to estimate the odds ratio and show that the estimator is strongly consistent and asymptotically normally distributed. Simulation studies indicate that the estimators perform satisfactorily in finite samples and against violations of assumptions. The proposed method is applied to a Sorafenib accessibility study of patients with advanced hepatocellular carcinoma.

REFERENCES

BRESLOW, N. E. and DAY, N. E. (1999). *Statistical Methods in Cancer Research*: *Volume* 1—*The Analysis of Case-Control Studies*. International Agency for Research on Cancer, Lyons.

FERGUSON, T. S. (1996). *A Course in Large Sample Theory*. *Texts in Statistical Science Series*. Chapman & Hall, London. MR1699953

GOLDSTEIN, L. and LANGHOLZ, B. (1992). Asymptotic theory for nested case-control sampling in the Cox regression model. *Ann*. *Statist*. **20** 1903–1928. MR1193318

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1952). *Inequalities*, 2nd ed. Cambridge Univ. Press, New York. MR0046395

KEOGH, R. H. and COX, D. R. (2014). *Case-Control Studies*. *Institute of Mathematical Statistics* (*IMS*) *Monographs* **4**. Cambridge Univ. Press, Cambridge. MR3443808

KUPPER, L. L., MCMICHAEL, A. J. and SPIRTAS, R. (1975). A hybrid epidemiology study design useful in estimating relative risk. *J*. *Amer*. *Statist*. *Assoc*. **99** 832–844.

LIDDELL, F. D. K., MCDONALD, J. C., THOMAS, D. C. and CUNLIFFE, S. V. (1977). Methods of cohort analysis: Appraisal by application to asbestos mining. *J*. *Roy*. *Statist*. *Soc*. *Ser*. *A* **140** 469–491.

MIETTINEN, O. S. (1976). Estimability and estimation in case-referent studies. *Am*. *J*. *Epidemiol*. **103** 226–235.

NURMINEN, M. (1989). Analysis of epidemiologic case-base studies for binary data. *Stat*. *Med*. **8** 1241–1254.

Key words and phrases. Case-only design, criminological epidemiology, disease registry, case-exposure association, imputation, pseudo-likelihood.

PRENTICE, R. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73** 1–11.

PRENTICE, R. and BRESLOW, N. E. (1978). Retrospective studies and failure time models. *Biometrika* **65** 153–158.

PRENTICE, R. L. and PYKE, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66** 403–411. MR0556730

ROBERTS, L. R. (2008). Sorafenib in liver cancer—Just the begining. *N. Engl. J. Med.* **359** 420–422.

SELF, S. G. and PRENTICE, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Statist.* **16** 64–81. MR0924857

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2254–2273
DOI: 10.1214/16-AOAS973

A STATISTICAL MODEL TO ASSESS (ALLELE-SPECIFIC) ASSOCIATIONS BETWEEN GENE EXPRESSION AND EPIGENETIC FEATURES USING SEQUENCING DATA

BY NAIM U. RASHID[*], WEI SUN[†] AND JOSEPH G. IBRAHIM[*]

University of North Carolina at Chapel Hill[*] *and*
Fred Hutchinson Cancer Research Center[†]

Sequencing techniques have been widely used to assess gene expression (i.e., RNA-seq) or the presence of epigenetic features (e.g., DNase-seq to identify open chromatin regions). In contrast to traditional microarray platforms, sequencing data are typically summarized in the form of discrete counts, and they are able to delineate allele-specific signals, which are not available from microarrays. The presence of epigenetic features are often associated with gene expression, both of which have been shown to be affected by DNA polymorphisms. However, joint models with the flexibility to assess interactions between gene expression, epigenetic features and DNA polymorphisms are currently lacking. In this paper, we develop a statistical model to assess the associations between gene expression and epigenetic features using sequencing data, while explicitly modeling the effects of DNA polymorphisms in either an allele-specific or nonallele-specific manner. We show that in doing so we provide the flexibility to detect associations between gene expression and epigenetic features, as well as conditional associations given DNA polymorphisms. We evaluate the performance of our method using simulations and apply our method to study the association between gene expression and the presence of DNase I Hypersensitive sites (DHSs) in HapMap individuals. Our model can be generalized to exploring the relationships between DNA polymorphisms and any two types of sequencing experiments, a useful feature as the variety of sequencing experiments continue to expand.

REFERENCES

1000 GENOMES PROJECT CONSORTIUM, ABECASIS, G. R., AUTON, A., BROOKS, L. D., DEPRISTO, M. A., DURBIN, R. M., HANDSAKER, R. E., KANG, H. M., MARTH, G. T. and MCVEAN, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* **491** 56–65.

AITCHISON, J. and HO, C.-H. (1989). The multivariate Poisson-log normal distribution. *Biometrika* **76** 643–653. MR1041409

BULMER, M. G. (1974). On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics* 101–110.

COWPER-SAL, R., ZHANG, X., WRIGHT, J. B., BAILEY, S. D., COLE, M. D., EECKHOUTE, J., MOORE, J. H., LUPIEN, M. et al. (2012). Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat. Genet.* **44** 1191–1198.

Key words and phrases. Bivariate binomial logistic-normal (BBLN) distribution, bivariate Poisson log-normal (BPLN) distribution, DNase-seq, genetics, genomics, RNA-seq.

DABNEY, A. and STOREY, J. D. (2015). qvalue: Q-value estimation for false discovery rate control. R package Version 1.38.0.

DANAHER, P. J. and HARDIE, B. G. S. (2005). Bacon with your eggs? Applications of a new bivariate beta-binomial distribution. *Amer. Statist.* **59** 282–286. MR2196349

DEGNER, J. F., PAI, A. A., PIQUE-REGI, R., VEYRIERAS, J. B., GAFFNEY, D. J., PICKRELL, J. K., DE LEON, S., MICHELINI, K., LEWELLEN, N., CRAWFORD, G. E. et al. (2012). DNaseI sensitivity QTLs are a major determinant of human expression variation. *Nature* **482** 390–394.

DJEBALI, S., DAVIS, C. A., MERKEL, A., DOBIN, A., LASSMANN, T., MORTAZAVI, A., TANZER, A., LAGARDE, J., LIN, W., SCHLESINGER, F. et al. (2012). Landscape of transcription in human cells. *Nature* **489** 101–108.

FAMOYE, F. (2010). On the bivariate negative binomial regression model. *J. Appl. Stat.* **37** 969–981. MR2757107

FANG, F., HODGES, E., MOLARO, A., DEAN, M., HANNON, G. J. and SMITH, A. D. (2012). Genomic landscape of human allele-specific DNA methylation. *Proc. Natl. Acad. Sci. USA* **109** 7332–7337.

GALLOPIN, M., RAU, A., JAFFRÉZIC, F. and CHEN, L. (2013). A hierarchical Poisson log-normal model for network inference from rna sequencing data. *PLoS ONE* **8**.

HARTZEL, J., AGRESTI, A. and CAFFO, B. (2001). Multinomial logit random effects models. *Stat. Model.* **1** 81–102.

HEINTZMAN, N. D., HON, G. C., HAWKINS, R. D., KHERADPOUR, P., STARK, A., HARP, L. F., YE, Z., LEE, L. K., STUART, R. K., CHING, C. W., CHING, K. A., ANTOSIEWICZ-BOURGET, J. E., LIU, H., ZHANG, X., GREEN, R. D., LOBANENKOV, V. V., STEWART, R., THOMSON, J. A., CRAWFORD, G. E., KELLIS, M. and REN, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459** 108–12.

JAENISCH, R. and BIRD, A. (2003). Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33 Suppl** 245–254.

LI, Y., WILLER, C. J., DING, J., SCHEET, P. and ABECASIS, G. R. (2010). MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34** 816–834.

LIU, Q. and PIERCE, D. A. (1994). A note on Gauss-Hermite quadrature. *Biometrika* **81** 624–629. MR1311107

MA, J., KOCKELMAN, K. M. and DAMIEN, P. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Anal. Prev.* **40** 964–975.

MAVROMMATIS, E., ARSLAN, A. D., SASSANO, A., HUA, Y., KROCZYNSKA, B. and PLATANIAS, L. C. (2013). Expression and regulatory effects of murine Schlafen (Slfn) genes in malignant melanoma and renal cell carcinoma. *J. Biol. Chem.* **288** 33006–33015.

MCDANIELL, R., LEE, B.-K., SONG, L., LIU, Z., BOYLE, A. P., ERDOS, M. R., SCOTT, L. J., MORKEN, M. A., KUCERA, K. S., BATTENHOUSE, A. et al. (2010). Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328** 235–239.

NYHOLT, D. R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **74** 765–769.

PARK, E. and LORD, D. (2007). Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. *Transp. Res. Rec.* **2019** 1–6.

PICKRELL, J. K., MARIONI, J. C., PAI, A. A., DEGNER, J. F., ENGELHARDT, B. E., NKADORI, E., VEYRIERAS, J.-B., STEPHENS, M., GILAD, Y. and PRITCHARD, J. K. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464** 768–772.

QUINLAN, A. R. and HALL, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26** 841–842.

RASHID, N. U., SUN, W. and IBRAHIM, J. G. (2016). Supplement to "A statistical model to assess (allele-specific) associations between gene expression and epigenetic features using sequencing data." DOI:10.1214/16-AOAS973SUPP.

ROZOWSKY, J., ABYZOV, A., WANG, J., ALVES, P., RAHA, D., HARMANCI, A., LENG, J., BJORNSON, R., KONG, Y., KITABAYASHI, N. et al. (2011). AlleleSeq: Analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* **7**.

SONG, L., ZHANG, Z., GRASFEDER, L. L., BOYLE, A. P., GIRESI, P. G., LEE, B. K., SHEFFIELD, N. C., GRÄF, S., HUSS, M., KEEFE, D. et al. (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* **21** 1757–1767.

SUN, W. (2012). A statistical framework for eQTL mapping using RNA-seq data. *Biometrics* **68** 1–11. MR2909848

SUN, W., YU, T. and LI, K.-C. (2007). Detection of eQTL modules mediated by activity levels of transcription factors. *Bioinformatics* **23** 2290–2297.

SUN, W., LIU, Y., CROWLEY, J. J., CHEN, T. H., ZHOU, H., CHU, H., HUANG, S., KUAN, P. F., LI, Y., MILLER, D., SHAW, G., WU, Y., ZHABOTYNSKY, V., MCMILLAN, L., ZOU, F., SULLIVAN, P. F. and PARDO-MANUEL DE VILLENA, F. (2015). IsoDOT detects differential RNA-isoform usage with respect to a categorical or continuous covariate with high sensitivity and specificity. *J. Amer. Statist. Assoc.* **110** 975–986.

THURMAN, R. E., RYNES, E., HUMBERT, R., VIERSTRA, J., MAURANO, M. T., HAUGEN, E., SHEFFIELD, N. C., STERGACHIS, A. B., WANG, H., VERNOT, B. et al. (2012). The accessible chromatin landscape of the human genome. *Nature* **489** 75–82.

TRAPNELL, C., PACHTER, L. and SALZBERG, S. L. (2009). TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics* **25** 1105–1111.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2274–2302
DOI: 10.1214/16-AOAS979

IMPROVING ICE SHEET MODEL CALIBRATION USING PALEOCLIMATE AND MODERN DATA

BY WON CHANG[1,2,*], MURALI HARAN[1,3,†], PATRICK APPLEGATE[1,3,†] AND DAVID POLLARD[1,3,4,†]

University of Cincinnati and Pennsylvania State University†*

Human-induced climate change may cause significant ice volume loss from the West Antarctic Ice Sheet (WAIS). Projections of ice volume change from ice sheet models and corresponding future sea-level rise have large uncertainties due to poorly constrained input parameters. In most future applications to date, model calibration has utilized only modern or recent (decadal) observations, leaving input parameters that control the long-term behavior of WAIS largely unconstrained. Many paleo-observations are in the form of localized time series, while modern observations are non-Gaussian spatial data; combining information across these types poses nontrivial statistical challenges. Here we introduce a computationally efficient calibration approach that utilizes both modern and paleo-observations to generate better constrained ice volume projections. Using fast emulators built upon principal component analysis and a reduced dimension calibration model, we can efficiently handle high-dimensional and non-Gaussian data. We apply our calibration approach to the PSU3D-ICE model which can realistically simulate long-term behavior of WAIS. Our results show that using paleo-observations in calibration significantly reduces parametric uncertainty, resulting in sharper projections about the future state of WAIS. One benefit of using paleo-observations is found to be that unrealistic simulations with overshoots in past ice retreat and projected future regrowth are eliminated.

REFERENCES

APPLEGATE, P. J., KIRCHNER, N., STONE, E. J., KELLER, K. and GREVE, R. (2012). An assessment of key model parametric uncertainties in projections of Greenland ice sheet behavior. *Cryosphere* **6** 589–606.

BAYARRI, M. J., BERGER, J. O., CAFEO, J., GARCIA-DONATO, G., LIU, F., PALOMO, J., PARTHASARATHY, R. J., PAULO, R., SACKS, J. and WALSH, D. (2007). Computer model validation with functional output. *Ann. Statist.* **35** 1874–1906. MR2363956

BHAT, K. S., HARAN, M. and GOES, M. (2010). Computer model calibration with multivariate spatial output: A case study. In *Frontiers of Statistical Decision Making and Bayesian Analysis* (M. H. Chen, P. Müller, D. Sun, K. Ye and D. K. Dey, eds.) 168–184. Springer, New York.

BHAT, K. S., HARAN, M., OLSON, R. and KELLER, K. (2012). Inferring likelihoods and climate system characteristics from climate models and multiple tracers. *Environmetrics* **23** 345–362. MR2935569

BINDSCHADLER, R. A., NOWICKI, S., ABE-OUCHI, A., ASCHWANDEN, A., CHOI, H., FASTOOK, J., GRANZOW, G., GREVE, R., GUTOWSKI, G., HERZFELD, U., JACKSON, C.,

Key words and phrases. Paleoclimate, West Antarctic Ice Sheet, computer model calibration, Gaussian process, dimension reduction.

JOHNSON, J., KHROULEV, C., LEVERMANN, A., LIPSCOMB, W. H., MARTIN, M. A., MORLIGHEM, M., PARIZEK, B. R., POLLARD, D., PRICE, S. F., REN, D., SAITO, F., SATO, T., SEDDIK, H., SEROUSSI, H., TAKAHASHI, K., WALKER, R. and WANG, W. L. (2013). Ice-sheet model sensitivities to environmental forcing and their use in projecting future sea level (the SeaRISE project). *J. Glaciol.* **59** 195–224.

BRIGGS, R., POLLARD, D. and TARASOV, L. (2013). A glacial systems model configured for large ensemble analysis of Antarctic deglaciation. *Cryosphere* **7** 1533–1589.

BRIGGS, R. D., POLLARD, D. and TARASOV, L. (2014). A data-constrained large ensemble analysis of Antarctic evolution since the Eemian. *Quat. Sci. Rev.* **103** 91–115.

BRIGGS, R. D. and TARASOV, L. (2013). How to evaluate model-derived deglaciation chronologies: A case study using Antarctica. *Quat. Sci. Rev.* **63** 109–127.

BRYNJARSDÓTTIR, J. and O'HAGAN, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse Probl.* **30** 114007, 24. MR3274591

CHANG, W., HARAN, M., OLSON, R. and KELLER, K. (2014a). Fast dimension-reduced climate model calibration and the effect of data aggregation. *Ann. Appl. Stat.* **8** 649–673. MR3262529

CHANG, W., APPLEGATE, P., HARAN, H. and KELLER, K. (2014b). Probabilistic calibration of a Greenland ice sheet model using spatially-resolved synthetic observations: Toward projections of ice mass loss with uncertainties. *Geosci. Model Dev.* **7** 1933–1943.

CHANG, W., HARAN, M., APPLEGATE, P. and POLLARD, D. (2016). Supplement to "Improving ice sheet model calibration using paleoclimate and modern data." DOI:10.1214/16-AOAS979SUPP.

CHANG, W., HARAN, M., APPLEGATE, P. and POLLARD, D. (2016). Calibrating an ice sheet model using high-dimensional binary spatial data. *J. Amer. Statist. Assoc.* **111** 57–72. MR3494638

CORNFORD, S. L., MARTIN, D. F., PAYNE, A. J., NG, E. G., LE BROCQ, A. M., GLADSTONE, R. M., EDWARDS, T. L., SHANNON, S. R., AGOSTA, C., VAN DEN BROEKE, M. R., HELLMER, H. H., KRINNER, G., LIGTENBERG, S. R. M., TIMMERMANN, R. and VAUGHAN, D. G. (2015). Century-scale simulations of the response of the West Antarctic Ice Sheet to a warming climate. *Cryosphere* **9** 1579–1600.

FAVIER, L., DURAND, G., CORNFORD, S. L., GUDMUNDSSON, G. H., GAGLIARDINI, O., GILLET-CHAULET, F., ZWINGER, T., PAYNE, A. J. and LE BROCQ, A. M. (2014). Retreat of Pine Island Glacier controlled by marine ice-sheet instability. *Nature Climate Change* **4** 171–121.

FELDMANN, J. and LEVERMANN, A. (2015). Collapse of the West Antarctic Ice Sheet after local destabilization of the Amundsen Basin. *Proc. Natl. Acad. Sci. USA* **112** 14191–14196.

FLEGAL, J. M., HARAN, M. and JONES, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statist. Sci.* **23** 250–260. MR2516823

FRETWELL, P., PRITCHARD, H. D., VAUGHAN, D. G., BAMBER, J. L., BARRAND, N. E., BELL, R., BIANCHI, C., BINGHAM, R. G., BLANKENSHIP, D. D., CASASSA, G., CATANIA, G., CALLENS, D., CONWAY, H., COOK, A. J., CORR, H. F. J., DAMASKE, D., DAMM, V., FERRACCIOLI, F., FORSBERG, R., FUJITA, S., GIM, Y., GOGINENI, P., GRIGGS, J. A., HINDMARSH, R. C. A., HOLMLUND, P., HOLT, J. W., JACOBEL, R. W., JENKINS, A., JOKAT, W., JORDAN, T., KING, E. C., KOHLER, J., KRABILL, W., RIGER-KUSK, M., LANGLEY, K. A., LEITCHENKOV, G., LEUSCHEN, C., LUYENDYK, B. P., MATSUOKA, K., MOUGINOT, J., NITSCHE, F. O., NOGI, Y., NOST, O. A., POPOV, S. V., RIGNOT, E., RIPPIN, D. M., RIVERA, A., ROBERTS, J., ROSS, N., SIEGERT, M. J., SMITH, A. M., STEINHAGE, D., STUDINGER, M., SUN, B., TINTO, B. K., WELCH, B. C., WILSON, D., YOUNG, D. A., XIANGBIN, C. and ZIRIZZOTTI, A. (2013). Bedmap2: Improved ice bed, surface and thickness datasets for Antarctica. *Cryosphere* **7** 375–393.

GLADSTONE, R. M., LEE, V., ROUGIER, J., PAYNE, A. J., HELLMER, H., LE BROCQ, A., SHEPHERD, A., EDWARDS, T. L., GREGORY, J. and CORNFORD, S. L. (2012). Calibrated prediction of Pine Island Glacier retreat during the 21st and 22nd centuries with a coupled flowline model. *Earth Planet. Sci. Lett.* **333** 191–199.

GOLLEDGE, N. R., MENVIEL, L., CARTER, L., FOGWILL, C. J., ENGLAND, M. H., CORTESE, G. and LEVY, R. H. (2014). Antarctic contribution to meltwater pulse 1A from reduced Southern Ocean overturning. *Nature Comm.* **5**.

GOLLEDGE, N. R., KOWALEWSKI, D. E., NAISH, T. R., LEVY, R. H., FOGWILL, C. J. and GASSON, E. G. W. (2015). The multi-millennial Antarctic commitment to future sea-level rise. *Nature* **526** 421–425.

GOMEZ, N., POLLARD, D. and HOLLAND, D. (2015). Sea-level feedback lowers projections of future Antarctic ice-sheet mass loss. *Nature Comm.* **6**.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*, 2nd ed. Springer, New York. MR2722294

HELLMER, H. H., KAUKER, F., TIMMERMANN, R., DETERMANN, J. and RAE, J. (2012). Twenty-first-century warming of a large Antarctic ice-shelf cavity by a redirected coastal current. *Nature* **485** 225–228.

HIGDON, D., GATTIKER, J., WILLIAMS, B. and RIGHTLEY, M. (2008). Computer model calibration using high-dimensional output. *J. Amer. Statist. Assoc.* **103** 570–583. MR2523994

JONES, G. L., HARAN, M., CAFFO, B. S. and NEATH, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101** 1537–1547. MR2279478

JOUGHIN, I., SMITH, B. E. and MEDLEY, B. (2014). Marine ice sheet collapse potentially under way for the Thwaites Glacier Basin, West Antarctica. *Science* **344** 735–738.

KENNEDY, M. C. and O'HAGAN, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 425–464. MR1858398

KIRSHNER, A. E., ANDERSON, J. B., JAKOBSSON, M., O'REGAN, M., MAJEWSKI, W. and NITSCHE, F. O. (2012). Post-LGM deglaciation in Pine Island Bay, West Antarctica. *Quat. Sci. Rev.* **38** 11–26.

LARTER, R. D., ANDERSON, J. B., GRAHAM, A. G., GOHL, K., HILLENBRAND, C.-D., JAKOBSSON, M., JOHNSON, J. S., KUHN, G., NITSCHE, F. O. and SMITH, J. A. (2014). Reconstruction of changes in the Amundsen Sea and Bellingshausen sea sector of the West Antarctic Ice Sheet since the last glacial maximum. *Quat. Sci. Rev.* **100** 55–86.

LEMPERT, R., SRIVER, R. L. and KELLER, K. (2012). *Characterizing uncertain sea level rise projections to support investment decisions*. California Energy Commission. Publication Number: CEC-500-2012-056.

LIU, Z., OTTO-BLIESNER, B. L., HE, F., BRADY, E. C., TOMAS, R., CLARK, P. U., CARLSON, A. E., LYNCH-STIEGLITZ, J., CURRY, W., BROOK, E., ERICKSON, D., JACOB, R., KUTZBACH, J. and CHENG, J. (2009). Transient simulation of last deglaciation with a new mechanism for Bølling–Allerød warming. *Science* **325** 310–314.

MARIS, M. N. A., VAN WESSEM, J. M., VAN DE BERG, W. J., DE BOER, B. and OERLEMANS, J. (2015). A model study of the effect of climate and sea-level change on the evolution of the Antarctic Ice Sheet from the Last Glacial Maximum to 2100. *Clim. Dynam.* **45** 837–851.

MCNEALL, D. J., CHALLENOR, P. G., GATTIKER, J. R. and STONE, E. J. (2013). The potential of an observational data set for calibration of a computationally expensive computer model. *Geosci. Model Dev.* **6** 1715–1728.

POLLARD, D. and DECONTO, R. M. (2009). Modelling West Antarctic Ice Sheet growth and collapse through the past five million years. *Nature* **458** 329–332.

POLLARD, D. and DECONTO, R. M. (2012a). A simple inverse method for the distribution of basal sliding coefficients under ice sheets, applied to Antarctica. *Cryosphere* **6** 1405–1444.

POLLARD, D. and DECONTO, R. M. (2012b). Description of a hybrid ice sheet-shelf model, and application to Antarctica. *Geosci. Model Dev.* **5** 1273–1295.

PRITCHARD, H. D., LIGTENBERG, S. R. M., FRICKER, H. A., VAUGHAN, D. G., VAN DEN BROEKE, M. R. and PADMAN, L. (2012). Antarctic ice-sheet loss driven by basal melting of ice shelves. *Nature* **484** 502–505.

RAISED CONSORTIUM (2014). A community-based geological reconstruction of Antarctic Ice Sheet deglaciation since the Last Glacial Maximum. *Quat. Sci. Rev.* **100** 1–9.

RITZ, C., EDWARDS, T. L., DURAND, G., PAYNE, A. J., PEYAUD, V. and HINDMARSH, R. C. (2015). Potential sea-level rise from Antarctic ice-sheet instability constrained by observations. *Nature* **528** 115–118.

SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4** 409–435. MR1041765

STONE, E. J., LUNT, D. J., RUTT, I. C. and HANNA, E. (2010). Investigating the sensitivity of numerical model simulations of the modern state of the Greenland ice-sheet and its future response to climate change. *Cryosphere* **4** 397–417.

WHITEHOUSE, P. L., BENTLEY, M. J. and LE BROCQ, A. M. (2012). A deglacial model for Antarctica: Geological constraints and glaciological modeling as a basis for a new model of Antarctic glacial isostatic adjustment. *Quat. Sci. Rev.* **32** 1–24.

WHITEHOUSE, P. L., BENTLEY, M. J., MILNE, G. A., KING, M. A. and THOMAS, I. D. (2012). A new glacial isostatic model for Antarctica: Calibrated and tested using observations of relative sea-level change and present-day uplifts. *Geophysical Journal International* **190** 1464–1482.

WINKELMANN, R., LEVERMANN, A., RIDGWELL, A. and CALDEIRA, K. (2015). Combustion of available fossil fuel resources sufficient to eliminate the Antarctic Ice Sheet. *Sci. Adv.* **1** e1500589.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2303–2324
DOI: 10.1214/16-AOAS980
© Institute of Mathematical Statistics, 2016

BAYESIAN INFERENCE FOR THE BROWN–RESNICK PROCESS, WITH AN APPLICATION TO EXTREME LOW TEMPERATURES[1]

BY EMERIC THIBAUD*, JUHA AALTO†, DANIEL S. COOLEY*, ANTHONY C. DAVISON‡ AND JUHA HEIKKINEN§

*Colorado State University**, *Finnish Meteorological Institute*†, *Ecole Polytechnique Fédérale de Lausanne*‡ *and Natural Resources Institute Finland*§

The Brown–Resnick max-stable process has proven to be well suited for modeling extremes of complex environmental processes, but in many applications its likelihood function is intractable and inference must be based on a composite likelihood, thereby preventing the use of classical Bayesian techniques. In this paper we exploit a case in which the full likelihood of a Brown–Resnick process can be calculated, using componentwise maxima and their partitions in terms of individual events, and we propose two new approaches to inference. The first estimates the partitions using declustering, while the second uses random partitions in a Markov chain Monte Carlo algorithm. We use these approaches to construct a Bayesian hierarchical model for extreme low temperatures in northern Fennoscandia.

REFERENCES

AALTO, J., LE ROUX, P. C. and LUOTO, M. (2014). The meso-scale drivers of temperature extremes in high-latitude Fennoscandia. *Climate Dynamics* **42** 237–252.

ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648

ASADI, P., DAVISON, A. C. and ENGELKE, S. (2015). Extremes on river networks. *Ann. Appl. Stat.* **9** 2023–2050. MR3456363

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, Chichester. With contributions from Daniel De Waal and Chris Ferro. MR2108013

BROWN, B. M. and RESNICK, S. I. (1977). Extreme values of independent stochastic processes. *J. Appl. Probab.* **14** 732–739. MR0517438

BUHL, S. and KLÜPPELBERG, C. (2016). Anisotropic Brown–Resnick space-time processes: Estimation and model assessment. *Extremes* **19** 627–660. MR3558348

CASSON, E. and COLES, S. (1999). Spatial regression models for extremes. *Extremes* **1** 449–468.

CASTRUCCIO, S., HUSER, R. and GENTON, M. G. (2015). High-order composite likelihood inference for max-stable distributions and processes. *J. Comput. Graph. Statist.* To appear. DOI:10.1080/10618600.2015.1086656.

CHAVEZ-DEMOULIN, V. and DAVISON, A. C. (2012). Modelling time series extremes. *REVSTAT* **10** 109–133. MR2912373

COLES, S. G. and TAWN, J. A. (1991). Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B* **53** 377–392. MR1108334

Key words and phrases. Global warming, likelihood-based inference, max-stable process, nonstationary extremes, partition, space-time declustering.

COOLEY, D., NYCHKA, D. and NAVEAU, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *J. Amer. Statist. Assoc.* **102** 824–840. MR2411647

DAVISON, A. C. and GHOLAMREZAEE, M. M. (2012). Geostatistics of extremes. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **468** 581–608. MR2874052

DAVISON, A. C., HUSER, R. and THIBAUD, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Math. Geosci.* **45** 511–529. MR3079649

DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27** 161–186. MR2963980

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York. MR2234156

DIEKER, A. B. and MIKOSCH, T. (2015). Exact simulation of Brown–Resnick random fields at a finite number of locations. *Extremes* **18** 301–314. MR3351818

DOMBRY, C., ÉYI-MINKO, F. and RIBATET, M. (2013). Conditional simulation of max-stable processes. *Biometrika* **100** 111–124. MR3034327

ENGELKE, S., MALINOWSKI, A., KABLUCHKO, Z. and SCHLATHER, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 239–265. MR3299407

FUENTES, M., HENRY, J. and REICH, B. (2013). Nonparametric spatial models for extremes: Application to extreme temperature data. *Extremes* **16** 75–101. MR3020178

HUSER, R. and DAVISON, A. C. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika* **100** 511–518. MR3068451

HUSER, R. and GENTON, M. G. (2016). Non-stationary dependence structures for spatial extremes. *J. Agric. Biol. Environ. Stat.* **21** 470–491. MR3542082

IPCC (2013). Summary for policymakers. In *Climate Change* 2013: *The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (T. F. STOCKER, D. QIN, G. K. PLATTNER, M. TIGNOR, S. K. ALLEN, J. BOSCHUNG, A. NAUELS, Y. XIA, V. BEX and P. M. MIDGLEY, eds.) 3–29. Cambridge Univ. Press, New York.

KABLUCHKO, Z., SCHLATHER, M. and DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* **37** 2042–2065. MR2561440

NADARAJAH, S. (2001). Multivariate declustering techniques. *Environmetrics* **12** 357–365.

NIKOLOULOPOULOS, A. K., JOE, H. and LI, H. (2009). Extreme value properties of multivariate t copulas. *Extremes* **12** 129–148. MR2515644

OPITZ, T. (2013). Extremal t processes: Elliptical domain of attraction and a spectral representation. *J. Multivariate Anal.* **122** 409–413. MR3189331

PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.* **105** 263–277. MR2757202

RAND, W. M. (1971). Objective criteria for the evaluation of clustering methods. *J. Amer. Statist. Assoc.* **66** 846–850.

RIBATET, M. (2013). Spatial extremes: Max-stable processes at work. *J. SFdS* **154** 156–177. MR3120441

RIBATET, M. (2015). SpatialExtremes: Modelling spatial extremes. R package version 2.0-2.

RIBATET, M., COOLEY, D. and DAVISON, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statist. Sinica* **22** 813–845. MR2954363

SANG, H. and GELFAND, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environ. Ecol. Stat.* **16** 407–426. MR2749848

SANG, H. and GELFAND, A. E. (2010). Continuous spatial process models for spatial extreme values. *J. Agric. Biol. Environ. Stat.* **15** 49–65. MR2755384

SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5** 33–44. MR1947786

SHABY, B. A. (2014). The open-faced sandwich adjustment for MCMC using estimating functions. *J. Comput. Graph. Statist.* **23** 853–876. MR3224659

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

SMITH, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript, Univ. Surrey. Available at http://www.stat.unc.edu/postscript/rs/spatex.pdf.

STEPHENSON, A. G. (2009). High-dimensional parametric modelling of multivariate extreme events. *Aust. N. Z. J. Stat.* **51** 77–88. MR2504104

STEPHENSON, A. and TAWN, J. (2005). Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika* **92** 213–227. MR2158621

THIBAUD, E. and OPITZ, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika* **102** 855–870. MR3431558

THIBAUD, E., AALTO, J., COOLEY, D. S., DAVISON, A. C. and HEIKKINEN, J. (2016). Supplement to "Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures." DOI:10.1214/16-AOAS980SUPP.

VIRTANEN, T., NEUVONEN, S. and NIKULA, A. (1998). Modelling topoclimatic patterns of egg mortality of *Epirrita autumnata* (Lepidoptera: Geometridae) with a Geographical Information System: Predictions for current climate and warmer climate scenarios. *Journal of Applied Ecology* **35** 311–322.

WADSWORTH, J. L. (2015). On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions. *Biometrika* **102** 705–711. MR3394285

WADSWORTH, J. L. and TAWN, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika* **101** 1–15. MR3180654

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2325–2348
DOI: 10.1214/16-AOAS981

A LAG FUNCTIONAL LINEAR MODEL FOR PREDICTION OF MAGNETIZATION TRANSFER RATIO IN MULTIPLE SCLEROSIS LESIONS

BY GINA-MARIA POMANN[*,1,3], ANA-MARIA STAICU[†,2,3], EDGAR J. LOBATON[†], AMANDA F. MEJIA[3,5], BLAKE E. DEWEY[§], DANIEL S. REICH[§], ELIZABETH M. SWEENEY[¶,3,4] AND RUSSELL T. SHINOHARA[‖,3,6]

Duke University[*], North Carolina State University[†], Indiana University Bloomington[‡], National Institute of Neurological Disorders and Stroke[§], Rice University[¶] and University of Pennsylvania[‖]

We propose a lag functional linear model to predict a response using multiple functional predictors observed at discrete grids with noise. Two procedures are proposed to estimate the regression parameter functions: (1) an approach that ensures smoothness for each value of time using generalized cross-validation; and (2) a global smoothing approach using a restricted maximum likelihood framework. Numerical studies are presented to analyze predictive accuracy in many realistic scenarios. The methods are employed to estimate a magnetic resonance imaging (MRI)-based measure of tissue damage (the magnetization transfer ratio, or MTR) in multiple sclerosis (MS) lesions, a disease that causes damage to the myelin sheaths around axons in the central nervous system. Our method of estimation of MTR within lesions is useful retrospectively in research applications where MTR was not acquired, as well as in clinical practice settings where acquiring MTR is not currently part of the standard of care. The model facilitates the use of commonly acquired imaging modalities to estimate MTR within lesions, and outperforms cross-sectional models that do not account for temporal patterns of lesion development and repair.

REFERENCES

BARKHOF, F. (2002). The clinico-radiological paradox in multiple sclerosis revisited. *Curr. Opin. Neurol.* **15** 239–245.

BESSE, P. and RAMSAY, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika* **51** 285–311. MR0848110

BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications. Lecture Notes in Statistics* **149**. Springer, New York. MR1783138

BREX, P. A., CICCARELLI, O., O'RIORDAN, J. I., SAILER, M., THOMPSON, A. J. and MILLER, D. H. (2002). A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *N. Engl. J. Med.* **346** 158–164.

CHEN, J. T., KUHLMANN, T., JANSEN, G. H., COLLINS, D. L., ATKINS, H. L., FREEDMAN, M. S., O'CONNOR, P. W., ARNOLD, D. L., GROUP, C. M. S. et al. (2007). Voxel-based

Key words and phrases. Functional data analysis, functional linear model, magnetization transfer ratio, image analysis.

analysis of the evolution of magnetization transfer ratio to quantify remyelination and demyelination with histopathological validation in a multiple sclerosis lesion. *NeuroImage* **36** 1152–1158.

Chen, J. T., Collins, D. L., Atkins, H. L., Freedman, M. S. and Arnold, D. L. (2008). Magnetization transfer ratio evolution with demyelination and remyelination in multiple sclerosis lesions. *Ann. Neurol.* **63** 254–262.

Crainiceanu, C. M., Staicu, A.-M. and Di, C.-Z. (2009). Generalized multilevel functional regression. *J. Amer. Statist. Assoc.* **104** 1550–1561. MR2750578

Di, C.-Z., Crainiceanu, C. M., Caffo, B. S. and Punjabi, N. M. (2009). Multilevel functional principal component analysis. *Ann. Appl. Stat.* **3** 458–488. MR2668715

Ferraty, F., Vieu, P. and Viguier-Pla, S. (2007). Factor-based comparison of groups of curves. *Comput. Statist. Data Anal.* **51** 4903–4910. MR2364548

Goldsmith, J., Greven, S. and Crainiceanu, C. I. P. R. I. A. N. (2012). Corrected confidence bands for functional data using principal components. *Biometrics*.

Hall, P., Müller, H.-G. and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. MR2278365

Harezlak, J., Coull, B. A., Laird, N. M., Magari, S. R. and Christiani, D. C. (2007). Penalized solutions to functional regression problems. *Comput. Statist. Data Anal.* **51** 4911–4925. MR2364549

Hawkins, C. P., Munro, P. M. G., MacKenzie, F., Kesselring, J., Tofts, P. S., Du Boulay, E. P. G. H., Landon, D. N. and McDonald, W. I. (1990). Duration and selectivity of blood-brain barrier breakdown in chronic relapsing experimental allergic encephalomyelitis studied by gadolinium-DTPA and protein markers. *Brain* **113** 365–378.

He, G., Müller, H.-G., Wang, J.-L. and Yang, W. (2010). Functional linear regression via canonical analysis. *Bernoulli* **16** 705–729. MR2730645

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer, New York. MR2920735

Ivanescu, A. E., Staicu, A.-M., Scheipl, F. and Greven, S. (2015). Penalized function-on-function regression. *Comput. Statist.* **30** 539–568. MR3357075

Jog, A., Roy, S., Carass, A. and Prince, J. L. (2013a). Pulse sequence based multi-acquisition MR intensity normalization. In *SPIE Medical Imaging* 86692H–86692H. International Society for Optics and Photonics.

Jog, A., Roy, S., Carass, A. and Prince, J. L. (2013b). Magnetic resonance image synthesis through patch regression. In *IEEE* 10*th International Symposium on Biomedical Imaging* (*ISBI*) 350–353. IEEE, New York.

Kim, K., Şentürk, D. and Li, R. (2011). Recent history functional linear models for sparse longitudinal data. *J. Statist. Plann. Inference* **141** 1554–1566. MR2747924

Krivobokova, T. and Kauermann, G. (2007). A note on penalized spline smoothing with correlated errors. *J. Amer. Statist. Assoc.* **102** 1328–1337. MR2412553

Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *Canad. J. Statist.* **31** 115–128. MR2016223

McDonald, W. I., Compston, A., Edan, G., Goodkin, D., Hartung, H.-P., Lublin, F. D., McFarland, H. F., Paty, D. W., Polman, C. H., Reingold, S. C. et al. (2001). Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. *Ann. Neurol.* **50** 121–127.

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

Meier, D. S., Weiner, H. L. and Guttmann, C. R. G. (2007). Time-series modeling of multiple sclerosis disease activity: A promising window on disease progression and repair potential? *Neurotherapeutics* **4** 485–498.

MEJIA, A., SWEENEY, E. M., DEWEY, B., NAIR, G., SATI, P., SHEA, C., REICH, D. S. and SHINOHARA, R. T. (2016). Statistical estimation of T1 relaxation times using conventional magnetic resonance imaging. *NeuroImage* **133** 176–188.

MEYER, M. J., COULL, B. A., VERSACE, F., CINCIRIPINI, P. and MORRIS, J. S. (2015). Bayesian function-on-function regression for multilevel functional data. *Biometrics* **71** 563–574. MR3402592

MORRIS, J. S. (2015). Functional regression. *Annual Reviews of Statistics and Its Applications* **2** 321–359.

POLMAN, C. H., REINGOLD, S. C., EDAN, G., FILIPPI, M., HARTUNG, H.-P., KAPPOS, L., LUBLIN, F. D., METZ, L. M., MCFARLAND, H. F., O'CONNOR, P. W. et al. (2005). Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald criteria". *Ann. Neurol.* **58** 840–846.

POMANN, G.-M., SWEENEY, E. M., REICH, D. S., STAICU, A.-M. and SHINOHARA, R. T. (2015). Scan-stratified case-control sampling for modeling blood-brain barrier integrity in multiple sclerosis. *Stat. Med.* **34** 2872–2880. MR3375986

POMANN, G.-M., STAICU, A., LOBATON, E. J., MEJIA, A. F., DEWEY, B. E., REICH, D. S., SWEENEY, E. M.E. M. and SHINOHARA, R. T.R. T. (2016). Supplement to "A lag functional linear model for prediction of magnetization transfer ratio in multiple sclerosis lesions." DOI:10.1214/16-AOAS981SUPP.

RAMSAY, J. O. and DALZELL, C. J. (1991). Some tools for functional data analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **53** 539–572. MR1125714

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

REICH, D. S., WHITE, R., CORTESE, I. C., VUOLO, O., SHEA, C. D., COLLINS, T. L. and PETKAU, J. (2015). Sample-size calculations for short-term proof-of-concept studies of tissue protection and repair in multiple sclerosis lesions via conventional clinical imaging. *Mult. Scler.* **21** 1693–1704.

REISS, P. T. and OGDEN, R. T. (2009). Smoothing parameter selection for a class of semiparametric linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 505–523. MR2649608

RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **53** 233–243. MR1094283

ROY, S., CARASS, A. and PRINCE, J. (2011). A compressed sensing approach for MR tissue contrast synthesis. In *Information Processing in Medical Imaging* 371–383. Springer, Berlin.

ROY, S., CARASS, A. and PRINCE, J. L. (2013). Magnetic resonance image example-based contrast synthesis. *IEEE Trans. Med. Imag.* **32** 2348–2363.

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

SCHEIPL, F. and GREVEN, S. (2015). Identifiability in penalized function-on-function regression models. Technical report, Univ. of Munich.

SCHEIPL, F., STAICU, A.-M. and GREVEN, S. (2015). Functional additive mixed models. *J. Comput. Graph. Statist.* **24** 477–501. MR3357391

SCHMIERER, K., SCARAVILLI, F., ALTMANN, D. R., BARKER, G. J. and MILLER, D. H. (2004). Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. *Ann. Neurol.* **56** 407–415.

STANISWALIS, J. G. and LEE, J. J. (1998). Nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **93** 1403–1418. MR1666636

SUTTNER, L., MEJIA, A., DEWEY, B., SATI, P., REICH, D. S. and SHINOHARA, R. T. (2015). Statistical estimation of white matter microstructure from conventional MRI. UPenn Biostatistics Working Papers. Working Paper 44.

Sweeney, E. M., Shinohara, R. T., Shea, C. D., Reich, D. S. and Crainiceanu, C. M. (2013a). Automatic lesion incidence estimation and detection in multiple sclerosis using multisequence longitudinal MRI. *Am. J. Neuroradiol.* **34** 68–73.

Sweeney, E. M., Shinohara, R. T., Shiee, N., Mateen, F. J., Chudgar, A. A., Cuzzocreo, J. L., Calabresi, P. A., Pham, D. L., Reich, D. S. and Crainiceanu, C. M. (2013b). OASIS is automated statistical inference for segmentation, with applications to multiple sclerosis lesion segmentation in MRI. *NeuroImage*: *Clinical* **2** 402–413.

Sweeney, E. M., Shinohara, R. T., Dewey, B. E., Schindler, M. K., Muschelli, J., Reich, D. S., Crainiceanu, C. M. and Eloyan, A. (2015). Relating multi-sequence longitudinal intensity profiles and clinical covariates in new multiple sclerosis lesions. Preprint. Available at arXiv:1509.08359.

Van Den Elskamp, I. J., Lembcke, J., Dattola, V., Beckmann, K., Pohl, C., Hong, W., Sandbrink, R., Wagner, K., Knol, D. L., Uitdehaag, B. et al. (2008). Persistent T1 hypointensity as an MRI marker for treatment efficacy in multiple sclerosis. *Mult. Scler.* **14** 764–769.

Wood, S. N. (2006). *Generalized Additive Models*: *An Introduction with R*. Chapman & Hall, Boca Raton, FL. MR2206355

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 3–36. MR2797734

Wu, Y., Fan, J. and Müller, H.-G. (2010). Varying-coefficient functional linear regression. *Bernoulli* **16** 730–758. MR2730646

Yao, F., Müller, H.-G. and Wang, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

Yao, F., Müller, H.-G. and Wang, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106

Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35** 1052–1079. MR2341698

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2349–2376
DOI: 10.1214/16-AOAS982

BOOTSTRAP AGGREGATING CONTINUAL REASSESSMENT METHOD FOR DOSE FINDING IN DRUG-COMBINATION TRIALS[1]

BY RUITAO LIN AND GUOSHENG YIN

University of Hong Kong

Phase I drug-combination trials are becoming commonplace in oncology. Most of the current dose-finding designs aim to quantify the toxicity probability space using certain prespecified yet complicated models. These models need to characterize not only each individual drug's toxicity profile, but also their interaction effects, which often leads to multi-parameter models. We propose a novel Bayesian adaptive design for drug-combination trials based on a robust dimension-reduction method. We continuously update the order of dose combinations and reduce the two-dimensional searching space to a one-dimensional line based on the estimated order. As a result, the common approaches to single-agent dose finding, such as the continual reassessment method (CRM), can be applied to drug-combination trials. We further utilize the ensemble technique in machine learning, the so-called bootstrap aggregating (bagging) in conjunction with Bayesian model averaging, to enhance the efficiency and reduce the variability of the proposed method. We conduct extensive simulation studies to examine the operating characteristics of the proposed method under various scenarios. Compared with existing competitive designs, the bagging CRM demonstrates its precision and robustness in terms of pinning down the correct dose combination. We apply the proposed bagging CRM to two recent cancer clinical trials with combined drugs for dose finding.

REFERENCES

AHN, C. (1998). An evaluation of phase I cancer clinical trial designs. *Stat. Med.* **17** 1537–1549.

BENDELL, J. C., JONES, S. F., HART, L., SPIGEL, D. R., LANE, C. M., EARWOOD, C., INFANTE, J. R., BARTON, J. and BURRIS, H. A. (2015). A phase Ib study of linsitinib (OSI-906), a dual inhibitor of IGF-1R and IR tyrosine kinase, in combination with everolimus as treatment for patients with refractory metastatic colorectal cancer. *Invest. New Drugs* **33** 187–193.

BREIMAN, L. (1996). Bagging predictors. *Mach. Learn.* **24** 123–140.

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

BRIL, G., DYKSTRA, R., PILLERS, C. and ROBERTSON, T. (1984). Algorithm AS 206: Isotonic regression in two independent variables. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **33** 352–357.

CHEUNG, Y. K. (2011). *Dose Finding by the Continual Reassessment Method*. Chapman & Hall/CRC, Boca Raton, FL.

CLYDE, M. A. and LEE, H. K. H. (2001). Bagging and the Bayesian bootstrap. In *Artificial Intelligence and Statistics* (T. Richardson and T. Jaakkola, eds.) 169–174. Elsevier, New York.

CONAWAY, M. R., DUNBAR, S. and PEDDADA, S. D. (2004). Designs for single- or multiple-agent phase I trials. *Biometrics* **60** 661–669. MR2089441

Key words and phrases. Bagging, Bayesian model averaging, dimension reduction, drug combination, maximum tolerated dose.

FAN, S. K., VENOOK, A. P. and LU, Y. (2009). Design issues in dose-finding phase I trials for combinations of two agents. *J. Biopharm. Statist.* **19** 509–523. MR2668723

FREUND, Y. and SCHAPIRE, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci.* **55** 119–139. MR1473055

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2000). Additive logistic regression: A statistical view of boosting. *Ann. Statist.* **28** 337–407. With discussion and a rejoinder by the authors. MR1790002

GANDHI, L., BAHLEDA, R., TOLANEY, S. M., KWAK, E. L., CLEARY, J. M., PANDYA, S. S., HOLLEBECQUE, A., ABBAS, R., ANANTHAKRISHNAN, R., BERKENBLIT, A., KRYGOWSKI, M., LIANG, Y., TURNBULL, K. W., SHAPIRO, G. I. and SORIA, J.-C. (2014). Phase I study of neratinib in combination with temsirolimus in patients with human epidermal growth factor receptor 2-dependent and other solid tumors. *J. Clin. Oncol.* **32** 68–75.

HARRINGTON, J. A., WHEELER, G. M., SWEETING, M. J., MANDER, A. P. and JODRELL, D. I. (2013). Adaptive designs for dual-agent phase I dose-escalation studies. *Nat. Rev. Clin. Oncol.* **10** 277–288.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294

HEYD, J. M. and CARLIN, P. B. (1999). Adaptive design improvements in the continual reassessment method for phase I studies. *Stat. Med.* **18** 1307–1321.

HIRAKAWA, A., HAMADA, C. and MATSUI, S. (2013). A dose-finding approach based on shrunken predictive probability for combinations of two agents in phase I trials. *Stat. Med.* **32** 4515–4525. MR3118372

HIRAKAWA, A., WAGES, N. A., SATO, H. and MATSUI, S. (2015). A comparative study of adaptive dose-finding designs for phase I oncology trials of combination therapies. *Stat. Med.* **34** 3194–3213. MR3412626

HOETING, J. A., MADIGAN, D., RAFTERY, A. E. and VOLINSKY, C. T. (1999). Bayesian model averaging: A tutorial. *Statist. Sci.* **14** 382–417. With comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors. MR1765176

HOUEDE, N., THALL, P. F., NGUYEN, H., PAOLETTI, X. and KRAMAR, A. (2010). Utility-based optimization of combination therapy using ordinal toxicity and efficacy in phase I/II trials. *Biometrics* **66** 532–540. MR2758833

HUANG, X., BISWAS, S., OKI, Y., ISSA, J.-P. and BERRY, D. A. (2007). A parallel phase I/II clinical trial design for combination therapies. *Biometrics* **63** 429–436. MR2370801

IASONOS, A. and O'QUIGLEY, J. (2014). Adaptive dose-finding studies: A review of model-guided phase I clinical trials. *J. Clin. Oncol.* **32** 2505–2511.

ISAKOFF, S. J., WANG, D., CAMPONE, M., CALLES, A., LEIP, E., TURNBULL, K., BARDY-BOUXIN, N., DUVILLIÉ, L. and CALVO, E. (2014). Bosutinib plus capecitabine for selected advanced solid tumours: Results of a phase 1 dose-escalation study. *Br. J. Cancer* **111** 2058–2066.

IVANOVA, A. and WANG, K. (2004). A non-parametric approach to the design and analysis of two-dimensional dose-finding trials. *Stat. Med.* **23** 1861–1870.

KORN, E. L. and SIMON, R. (1991). Selecting dose-intense drug combinations: Metastatic breast cancer. *Breast Cancer Res. Treat.* **20** 155–166.

KORN, E. L. and SIMON, R. (1993). Using the tolerable-dose diagram in the design of phase I combination chemotherapy trials. *J. Clin. Oncol.* **11** 794–801.

KORN, E. L., MIDTHUNE, D., CHEN, T. T., RUBINSTEIN, L. V., CHRISTIAN, M. C. and SIMON, R. M. (1994). A comparison of two phase I trial designs. *Stat. Med.* **13** 1799–1806.

KRAMAR, A., LEBECQ, A. and CANDALH, E. (1999). Continual reassessment methods in phase I trials of the combination of two drugs in oncology. *Stat. Med.* **18** 1849–1864.

LIN, R. and YIN, G. (2016). Bayesian optimal interval design for dose finding in drug-combination trials. *Stat. Methods Med. Res.*, DOI:10.1177/0962280215594494.

MANDER, A. P. and SWEETING, M. J. (2015). A product of independent beta probabilities dose escalation design for dual-agent phase I trials. *Stat. Med.* **34** 1261–1276. MR3322767

MANDREKAR, S. J. (2014). Dose-finding trial designs for combination therapies in oncology. *J. Clin. Oncol.* **32** 65–67.

O'QUIGLEY, J. and CONAWAY, M. (2010). Continual reassessment and related dose-finding designs. *Statist. Sci.* **25** 202–216. MR2789990

O'QUIGLEY, J., PEPE, M. and FISHER, L. (1990). Continual reassessment method: A practical design for phase 1 clinical trials in cancer. *Biometrics* **46** 33–48. MR1059105

ORON, A. P. and HOFF, P. D. (2013). Small-sample behavior of novel phase I cancer trial designs. *Clin. Trials* **10** 63–80.

PAPADATOS-PASTOS, D., LUKEN, M. D. M. and YAP, T. A. (2015). Combining targeted therapeutics in the era of precision medicine. *Br. J. Cancer* **112** 1–3.

RAFTERY, A. E., MADIGAN, D. and HOETING, J. A. (1997). Bayesian model averaging for linear regression models. *J. Amer. Statist. Assoc.* **92** 179–191. MR1436107

RIVIERE, M.-K., DUBOIS, F. and ZOHAR, S. (2015a). Competing designs for drug combination in phase I dose-finding clinical trials. *Stat. Med.* **34** 1–12. MR3286233

RIVIERE, M.-K., YUAN, Y., DUBOIS, F. and ZOHAR, S. (2014). A Bayesian dose-finding design for drug combination clinical trials based on the logistic model. *Pharm. Stat.* **13** 247–257.

RIVIERE, M. K., LE TOURNEAU, C., PAOLETTI, X., DUBOIS, F. and ZOHAR, S. (2015b). Designs of drug-combination phase I trials in oncology: A systematic review of the literature. *Ann. Oncol.* **26** 669–674.

RUBIN, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.* **9** 130–134. MR0600538

SAURA, C., GARCIA-SAENZ, J. A., XU, B., HARB, W., MOROOSE, R., PLUARD, T., CORTES, J., KIGER, C., GERMA, C., WANG, K., MARTIN, M., BASELGA, J. and KIM, S. B. (2014). Safety and efficacy of neratinib in combination with capecitabine in patients with metastatic human epidermal growth factor receptor 2-positive breast cancer. *J. Clin. Oncol.* **32** 3626–3633.

SHEN, L. Z. and O'QUIGLEY, J. (1996). Consistency of continual reassessment method under model misspecification. *Biometrika* **83** 395–405. MR1439791

SIEGEL, D. S., RICHARDSON, P., DIMOPOULOS, M., MOREAU, P., MITSIADES, C., WEBER, D., HOUP, J., GAUSE, C., VUOCOLO, S., EID, J., GRAEF, T. and ANDERSON, K. C. (2014). Vorinostat in combination with lenalidomide and dexamethasone in patients with relapsed or refractory multiple myeloma. *Blood Cancer J.* **4** e182.

THALL, P. F. (2010). Bayesian models and decision algorithms for complex early phase clinical trials. *Statist. Sci.* **25** 227–244. MR2789992

THALL, P. F., MILLIKAN, R. E., MUELLER, P. and LEE, S.-J. (2003). Dose-finding with two agents in phase I oncology trials. *Biometrics* **59** 487–496. MR2004253

TIGHIOUART, M., PIANTADOSI, S. and ROGATKO, A. (2014). Dose finding with drug combinations in cancer phase I clinical trials using conditional escalation with overdose control. *Stat. Med.* **33** 3815–3829. MR3260662

ULLENHAG, G. J., ROSSMANN, E. and LILJEFORS, M. (2015). A phase I dose-escalation study of lenalidomide in combination with gemcitabine in patients with advanced pancreatic cancer. *PLoS ONE* **10** e0121197.

WAGES, N. A., CONAWAY, M. R. and O'QUIGLEY, J. (2011). Continual reassessment method for partial ordering. *Biometrics* **67** 1555–1563. MR2872406

WANG, K. and IVANOVA, A. (2005). Two-dimensional dose finding in discrete dose space. *Biometrics* **61** 217–222. MR2135863

WILKY, B. A., RUDEK, M. A., AHMED, S., LAHERU, D. A., COSGROVE, D., DONEHOWER, R. C., NELKIN, B., BALL, D., DOYLE, L. A., CHEN, H., YE, X., BIGLEY, G., WOMACK, C. and AZAD, N. S. (2015). A phase I trial of vertical inhibition of IGF signalling using cixutumumab, an anti-IGF-1R antibody, and selumetinib, an MEK 1/2 inhibitor, in advanced solid tumours. *Br. J. Cancer* **112** 24–31.

YIN, G. (2012). *Clinical Trial Design*: *Bayesian and Frequentist Adaptive Methods*. John Wiley & Sons, Hoboken, New Jersey.

YIN, G. and LIN, R. (2015). Comments on 'Competing designs for drug combination in phase I dose-finding clinical trials' by M.-K. Riviere, F. Dubois, and S. Zohar [MR3286233]. *Stat*. *Med*. **34** 13–17. MR3286234

YIN, G. and YUAN, Y. (2009a). Bayesian dose finding in oncology for drug combinations by copula regression. *J*. *R*. *Stat*. *Soc*. *Ser*. *C*. *Appl*. *Stat*. **58** 211–224. MR2649671

YIN, G. and YUAN, Y. (2009b). Bayesian model averaging continual reassessment method in phase I clinical trials. *J*. *Amer*. *Statist*. *Assoc*. **104** 954–968. MR2750228

YUAN, Z. and CHAPPELL, R. (2004). Isotonic designs for phase I cancer clinical trials with multiple risk groups. *Clin*. *Trials* **1** 499–508.

YUAN, Y. and YIN, G. (2008). Sequential continual reassessment method for two-dimensional dose finding. *Stat*. *Med*. **27** 5664–5678. MR2573775

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2377–2404
DOI: 10.1214/16-AOAS986

A PHYLOGENETIC LATENT FEATURE MODEL FOR CLONAL DECONVOLUTION[1]

BY FRANCESCO MARASS[*], FLORENT MOULIERE[*], KE YUAN[†], NITZAN ROSENFELD[*] AND FLORIAN MARKOWETZ[*]

University of Cambridge[*] *and University of Glasgow*[†]

Tumours develop in an evolutionary process, in which the accumulation of mutations produces subpopulations of cells with distinct mutational profiles, called clones. This process leads to the genetic heterogeneity widely observed in tumour sequencing data, but identifying the genotypes and frequencies of the different clones is still a major challenge. Here, we present Cloe, a phylogenetic latent feature model to deconvolute tumour sequencing data into a set of related genotypes. Our approach extends latent feature models by placing the features as nodes in a latent tree. The resulting model can capture both the acquisition and the loss of mutations, as well as episodes of convergent evolution. We establish the validity of Cloe on synthetic data and assess its performance on controlled biological data, comparing our reconstructions to those of several published state-of-the-art methods. We show that our method provides highly accurate reconstructions and identifies the number of clones, their genotypes and frequencies even at a modest sequencing depth. As a proof of concept, we apply our model to clinical data from three cases with chronic lymphocytic leukaemia and one case with acute myeloid leukaemia.

REFERENCES

APARICIO, S. and CALDAS, C. (2013). The implications of clonal genome evolution for cancer medicine. *N. Engl. J. Med.* **368** 842–851.

BEERENWINKEL, N., SCHWARZ, R. F., GERSTUNG, M. and MARKOWETZ, F. (2015). Cancer evolution: Mathematical models and computational inference. *Syst. Biol.* **64** e1–e25.

DESHWAR, A. G., VEMBU, S., YUNG, C. K., JANG, G. H., STEIN, L. and MORRIS, Q. (2015). PhyloWGS: Reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol.* **16** 35.

EL-KEBIR, M., OESPER, L., ACHESON-FIELD, H. and RAPHAEL, B. J. (2015). Reconstruction of clonal trees and tumor composition from multi-sample sequencing data. *Bioinform.* **31** i62–i70.

FEARON, E. R. and VOGELSTEIN, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* **61** 759–767.

FISCHER, A., VÁZQUEZ-GARCÍA, I., ILLINGWORTH, C. J. and MUSTONEN, V. (2014). High-definition reconstruction of clonal composition in cancer. *Cell Rep.* **7** 1740–1752.

FORSHEW, T., MURTAZA, M., PARKINSON, C., GALE, D., TSUI, D. W. Y., KAPER, F., DAWSON, S.-J., PISKORZ, A. M., JIMENEZ-LINAN, M., BENTLEY, D., HADFIELD, J., MAY, A. P., CALDAS, C., BRENTON, J. D. and ROSENFELD, N. (2012). Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl. Med.* **4** 136ra68.

Key words and phrases. Clonal deconvolution, tumour heterogeneity, latent feature model, phylogeny, admixture.

GERLINGER, M., ROWAN, A. J., HORSWELL, S., LARKIN, J., ENDESFELDER, D., GRONROOS, E., MARTINEZ, P., MATTHEWS, N., STEWART, A., TARPEY, P., VARELA, I., PHILLIMORE, B., BEGUM, S., MCDONALD, N. Q., BUTLER, A., JONES, D., RAINE, K., LATIMER, C., SANTOS, C. R., NOHADANI, M., EKLUND, A. C., SPENCER-DENE, B., CLARK, G., PICKERING, L., STAMP, G., GORE, M., SZALLASI, Z., DOWNWARD, J., FUTREAL, P. A. and SWANTON, C. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366** 883–892.

GEYER, C. J. (1991). Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics*: *Proceedings of the* 23*rd Symposium on the Interface* (E. M. Keramidas, ed.) 156–163. Interface Foundation of North America.

GHAHRAMANI, Z. and GRIFFITHS, T. L. (2005). Infinite latent feature models and the Indian buffet process. In *Advances in Neural Information Processing Systems* 475–482.

GRIFFITH, M., MILLER, C. A., GRIFFITH, O. L., KRYSIAK, K., SKIDMORE, Z. L., RAMU, A., WALKER, J. R., DANG, H. X., TRANI, L., LARSON, D. E., DEMETER, R. T., WENDL, M. C., MCMICHAEL, J. F., AUSTIN, R. E., MAGRINI, V., MCGRATH, S. D., LY, A., KULKARNI, S., CORDES, M. G., FRONICK, C. C., FULTON, R. S., MAHER, C. A., DING, L., KLCO, J. M., MARDIS, E. R., LEY, T. J. and WILSON, R. K. (2015). Optimizing cancer genome sequencing and analysis. *Cell Syst.* **1** 210–223.

HEAUKULANI, C., KNOWLES, D. A. and GHAHRAMANI, Z. (2014). Beta diffusion trees and hierarchical feature allocations. Preprint. Available at arXiv:1408.3378.

JI, Y. (2016). Biostatistics and Bioinformatics Lab—Software. Available at http://health.bsd.uchicago.edu/yji/soft.html. Accessed: 2016-02-05.

JIAO, W., VEMBU, S., DESHWAR, A. G., STEIN, L. and MORRIS, Q. (2014). Inferring clonal evolution of tumors from single nucleotide somatic mutations. *BMC Bioinform.* **15** 1.

MARASS, F., MOULIERE, F., YUAN, K., ROSENFELD, N. and MARKOWETZ, F. (2016). Supplement to "A phylogenetic latent feature model for clonal deconvolution." DOI:10.1214/16-AOAS986SUPPA, DOI:10.1214/16-AOAS986SUPPB.

MILLER, K. T., GRIFFITHS, T. and JORDAN, M. I. (2012). The phylogenetic Indian buffet process: A non-exchangeable nonparametric prior for latent features. Preprint. Available at arXiv:1206.3279.

MILLER, C. A., WHITE, B. S., DEES, N. D., GRIFFITH, M., WELCH, J. S., GRIFFITH, O. L., VIJ, R., TOMASSON, M. H., GRAUBERT, T. A., WALTER, M. J. et al. (2014). SciClone: Inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput. Biol.* **10** e1003665.

NIK-ZAINAL, S., LOO, P. V., WEDGE, D. C., ALEXANDROV, L. B., GREENMAN, C. D., LAU, K. W., RAINE, K., JONES, D., MARSHALL, J., RAMAKRISHNA, M., SHLIEN, A., COOKE, S. L., HINTON, J., MENZIES, A., STEBBINGS, L. A., LEROY, C., JIA, M., RANCE, R., MUDIE, L. J., GAMBLE, S. J., STEPHENS, P. J., MCLAREN, S., TARPEY, P. S., PAPAEMMANUIL, E., DAVIES, H. R., VARELA, I., MCBRIDE, D. J., BIGNELL, G. R., LEUNG, K., BUTLER, A. P., TEAGUE, J. W., MARTIN, S., JÖNSSON, G., MARIANI, O., BOYAULT, S., MIRON, P., FATIMA, A., LANGERØD, A., APARICIO, S. A. J. R., TUTT, A., SIEUWERTS, A. M., BORG, Å., THOMAS, G., SALOMON, A. V., RICHARDSON, A. L., BØRRESEN-DALE, A.-L., FUTREAL, P. A., STRATTON, M. R., CAMPBELL, P. J. and BREAST CANCER WORKING GROUP OF THE INTERNATIONAL CANCER GENOME CONSORTIUM (2012). The life history of 21 breast cancers. *Cell* **149** 994–1007.

NOWELL, P. C. (1976). The clonal evolution of tumor cell populations. *Science* **194** 23–28.

RONQUIST, F., HUELSENBECK, J. P. and TESLENKO, M. (2005). MrBayes version 3.2 Manual: Tutorials and Model Summaries. http://mrbayes.sourceforge.net/mb3.2_manual.pdf. [Online; accessed 29 June 2016].

ROTH, A., KHATTRA, J., YAP, D., WAN, A., LAKS, E., BIELE, J., HA, G., APARICIO, S., BOUCHARD-CÔTÉ, A. and SHAH, S. P. (2014). PyClone: Statistical inference of clonal population structure in cancer. *Nat. Methods* **11** 396–398.

SCHUH, A., BECQ, J., HUMPHRAY, S., ALEXA, A., BURNS, A., CLIFFORD, R., FELLER, S. M., GROCOCK, R., HENDERSON, S., KHREBTUKOVA, I. et al. (2012). Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* **120** 4191–4196.

SCHWARZ, R. F., NG, C. K., COOOKE, S. L., NEWMAN, S., TEMPLE, J., PISKORZ, A. M., GALE, D., SAYAL, K., MURTAZA, M., BALDWIN, P. J., ROSENFELD, N., EARL, H. M., SALA, E., JIMENEZ-LINAN, M., PARKINSON, C. A., MARKOWETZ, F. and BRENTON, J. D. (2015). Spatial and temporal heterogeneity in high-grade serous ovarian cancer: A phylogenetic reconstruction. *PLoS Med* **12** e1001789.

SENGUPTA, S., WANG, J., LEE, J., MÜLLER, P., GULUKOTA, K., BANERJEE, A. and JI, Y. (2015). BayClone: Bayesian nonparametric inference of tumor subclones using NGS data. In *Pacific Symposium on Biocomputing* **20** 467. World Scientific.

STRATTON, M. R., CAMPBELL, P. J. and FUTREAL, P. A. (2009). The cancer genome. *Nature* **458** 719–724.

YUAN, K., SAKOPARNIG, T., MARKOWETZ, F. and BEERENWINKEL, N. (2015). BitPhylogeny: A probabilistic framework for reconstructing intra-tumor phylogenies. *Genome Biol.* **16** 36.

ZARE, H., WANG, J., HU, A., WEBER, K., SMITH, J., NICKERSON, D., SONG, C., WITTEN, D., BLAU, C. A. and NOBLE, W. S. (2014). Inferring clonal composition from multiple sections of a breast cancer. *PLoS Comput. Biol.* **10** e1003703.

The Annals of Applied Statistics
2016, Vol. 10, No. 4, 2405–2426
DOI: 10.1214/16-AOAS988

EXPLOITING TIMSS AND PIRLS COMBINED DATA: MULTIVARIATE MULTILEVEL MODELLING OF STUDENT ACHIEVEMENT[1]

BY LEONARDO GRILLI[*], FULVIA PENNONI[†], CARLA RAMPICHINI[*] AND ISABELLA ROMEO[‡]

University of Florence,[*] University of Milano-Bicocca[†] and Mario Negri[‡]

We illustrate how to perform a multivariate multilevel analysis in the complex setting of large-scale assessment surveys, dealing with plausible values and accounting for the survey design. In particular, we consider the Italian sample of the TIMSS&PIRLS 2011 Combined International Database on fourth grade students. The multivariate approach jointly considers educational achievement in Reading, Mathematics and Science, thus allowing us to test for differential associations of the covariates with the three outcomes, and to estimate the residual correlations among pairs of outcomes within and between classes. Multilevel modelling allows us to disentangle student and contextual factors affecting achievement. We also account for territorial differences in wealth by means of an index from an external data source. The model residuals point out classes with high or low performance. As educational achievement is measured by plausible values, the estimates are obtained through multiple imputation formulas.

REFERENCES

AMMERMUELLER, A. and PISCHKE, J. S. (2009). Peer effects in European primary schools: Evidence from the progress in international reading literacy study. *J. Labor. Econ.* **27** 315–348.

ASPAROUHOV, T. (2006). General multi-level modeling with sampling weights. *Comm. Statist. Theory Methods* **35** 439–460. MR2274063

BARTOLUCCI, F., PENNONI, F. and VITTADINI, G. (2011). Assessment of school performance through a multilevel latent Markov Rasch model. *J. Educ. Behav. Stat.* **36** 491–522.

BOUHLILA, D. S. and SELLAOUTI, F. (2013). Multiple imputation using chained equations for missing data in TIMSS: A case study. *Large Scale Assess. Educ.* **1** 1–33.

CHIU, M. M. and XIHUA, Z. (2008). Family and motivation effects on mathematics achievement: Analyses of students in 41 countries. *Learn. Instr.* **18** 321–336.

FOX, J.-P. and GLAS, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika* **66** 271–288. MR1836937

FOY, P. (2013). TIMSS and PIRLS 2011 user guide for the fourth grade combined international database. TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill, MA. Available at http://timssandpirls.bc.edu/timsspirls2011/international-database.html.

FOY, P. and O'DWYER, L. M. (2013). Technical Appendix B. School effectiveness models and analyses. In *TIMSS and PIRLS* 2011 *Relationships Among Reading*, *Mathematics*, *and Science Achievement at the Fourth Grade-Implications for Early Learning* (M. O. Martin and V. S. Mullis,

Key words and phrases. Hierarchical linear model, large-scale assessment data, multiple imputation, plausible values, school effectiveness, secondary data analysis.

eds.). TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill, MA. Available at http://timssandpirls.bc.edu/timsspirls2011/downloads/TP11_Relationship_Report.pdf.

GOLDSTEIN, H. (2004). International comparisons of student attainment: Some issues arising from the PISA study. *Assessment in Education*: *Principles*, *Policy & Practice* **11** 319–330.

GOLDSTEIN, H. (2011). *Multilevel Statistical Models*, 4th ed. Wiley, New York.

GOLDSTEIN, H., CARPENTER, J. R. and BROWNE, W. J. (2014). Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. *J. Roy. Statist. Soc. Ser. A* **177** 553–564. MR3249673

GRADY, M. W. and BERETVAS, S. N. (2010). Incorporating student mobility in achievement growth modeling: A cross-classified multiple membership growth curve model. *Multivar. Behav. Res.* **45** 393–419.

GRILLI, L. and RAMPICHINI, C. (2015). Specification of random effects in multilevel models: A review. *Qual. Quant.* **49** 967–976.

HAMMOURI, H. A. M. (2004). Attitudinal and motivational variables related to mathematics achievement in Jordan: Findings from the third international mathematics and science study (TIMSS). *Educ. Res.* **46** 241–257.

HANUSHEK, E. A. and WOESSMANN, L. (2011). The economics of international differences in educational achievement. In *Handbook of the Economics of Education* (E. A. Hanushek, S. Machin and L. Woessmann, eds.) **3**. Elsevier, The Netherlands.

ISTITUTO TAGLIACARNE (2011). Reddito e occupazione nelle province Italiane dal 1861 ad oggi. Istituto Tagliacarne, Roma.

JERRIM, J. and MICKLEWRIGHT, J. (2014). Socio-economic gradients in children cognitive skills: Are cross-country comparisons robust to who reports family background? *Eur. Sociol. Rev.* **30** 766–781.

JOHNSON, M. S. and JENKINS, F. (2005). A Bayesian hierarchical model for large-scale educational surveys: An application to the international assessment of educational progress. ETS Research report RR-04-38. Educational Testing Service, Princeton, NJ.

JONCAS, M. and FOY, P. (2013). Sample design in TIMSS and PIRLS. In *Methods and Procedures in TIMSS and PIRLS* 2011 (M. O. Martin and I. V. S. Mullis, eds.). TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill, MA. Available at http://timssandpirls.bc.edu/methods/pdf/TP_Sampling_Design.pdf.

KIRSCH, I., DE JONG, J., LAFONTAINE, D., MCQUEEN, J., MENDELOVITS, J. and MONSEUR, C. (2002). *Reading for Change*. *Performance and Engagement Across Countries*. *Results from Pisa* 2000. OECD, Paris.

KREINER, S. and CHRISTENSEN, K. B. (2014). Analyses of model fit and robustness. A new look at the PISA scaling model underlying ranking of countries according to reading literacy. *Psychometrika* **79** 210–231. MR3255117

KREUTER, F., ECKMAN, S., MAAZ, K. and WATERMANN, R. (2010). Children's reports of parents'education level: Does it matter whom you ask and what you ask about. *Surv. Res. Meth.* **4** 127–138.

KYRIAKIDES, L. (2008). Testing the validity of the comprehensive model of educational effectiveness: A step towards the development of a dynamic model of effectiveness. *Sch. Eff. Sch. Improv.* **19** 429–446.

LADD, H. and WALSH, R. (2002). Implementing value-added measures of school effectiveness: Getting the incentives right. *Econ. Educ. Rev.* **21** 1–17.

LI, K. H., MENG, X.-L., RAGHUNATHAN, T. E. and RUBIN, D. B. (1991). Significance levels from repeated p-values with multiply-imputed data. *Statist. Sinica* **1** 65–92. MR1101316

MARTIN, M. O. and MULLIS, I. V. S. (2012). Methods and procedures in TIMSS and PIRLS 2011. TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill, MA.

MARTIN, M. O. and MULLIS, I. V. S. (2013). Timss and Pirls 2011: Relationships Among Reading, Mathematics, and Science Achievement at the Fourth Grade-Implications for Early Learning. TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill, MA.

MISLEVY, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika* **56** 177–196.

RABE-HESKETH, S. and SKRONDAL, A. (2006). Multilevel modelling of complex survey data. *J. Roy. Statist. Soc. Ser. A* **169** 805–827. MR2291345

RAUDENBUSH, S. W. and WILLMS, J. D. (1995). The estimation of school effects. *J. Educ. Behav. Stat.* **20** 307–335.

REEVE, J. and JANG, H. (2006). What teachers say and do to support students' autonomy during a learning activity. *J. Educ. Psychol.* **98** 209–218.

RUBIN, D. (2002). *Multiple Imputation for Nonresponse in Sample Surveys*. Wiley, New York.

RUTKOWSKI, L., VON DAVIER, M. and RUTKOWSKI, D. (2014). *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*. Chapnam & Hall, Boca Raton.

RUTKOWSKI, L., GONZALEZ, E., JONCAS, M. and VON DAVIER, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educ. Res.* **39** 142–151.

SANI, C. and GRILLI, L. (2011). Differential variability of test scores among schools: A multilevel analysis of the fifth-grade Invalsi test using heteroscedastic random effects. *J. Appl. Quant. Meth.* **6** 88–99.

SCHAFER, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Stat. Neerl.* **57** 19–35. MR2055519

SNIJDERS, T. A. B. and BOSKER, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd ed. Sage Publications, Los Angeles, CA. MR3137621

STATACORP (2013). Stata: Release 13. Statistical Software. StataCorp LP, College Station, TX.

STONGE, J. H., WARD, T. J. and GRANT, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *J. Teach. Educ.* **62** 339–355.

TEKWE, C., CARTER, R., MA, C., ALGINA, J., LUCAS, M., ROTH, J., ARIET, M., FISHER, T. and RESNICK, M. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *J. Educ. Behav. Stat.* **29** 11–36.

TRANMER, M. and STEEL, D. G. (2001). Ignoring a level in a multilevel model: Evidence from UK census data. *Environ. Plann. A* **33** 941–948.

VON DAVIER, M., GONZALEZ, E. and MISLEVY, R. (2009). What are plausible values and why are they useful? In *ERI Monograph Series: Issues and Methodologies in Large-Scale Assessments* (M. von Davier and D. Hastedt, eds.) **2** 9–36.

WANG, Z., OSTERLIND, S. and BERGIN, D. (2012). Building mathematics achievement models in four countries using TIMSS 2003. *Int. J. Sci. Math. Educ.* **10** 1215–1242.

WEIRICH, S., HAAG, N., HECHT, M., BÖHME, K., SIEGLE, T. and LÜDTKE, O. (2014). Nested multiple imputation in large-scale assessments. *Large Scale Assess. Educ.* **2** 1–18.

WU, M. (2005). The role of plausible values in large-scale surveys. *Stud. Educ. Eval.* **31** 114–128.

YANG, M., GOLDSTEIN, H., BROWNE, W. and WOODHOUSE, G. (2002). Multivariate multilevel analyses of examination results. *J. Roy. Statist. Soc. Ser. A* **165** 137–153. MR1909740