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. 18, No. 1—March 2024

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*

Correction

Vol. 18, No. 1—March 2024

THE ANNALS OF APPLIED STATISTICS Vol. 18, No. 1, pp. 1–903 March 2024

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

IMS OFFICERS

President: Michael Kosorok, Department of Biostatistics and Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, Chapel Hill, NC 27599, USA

President-Elect: Tony Cai, Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104, USA

Past President: Peter Bühlmann, Seminar für Statistik, ETH Zürich, 8092 Zürich, Switzerland

Executive Secretary: Peter Hoff, Department of Statistical Science, Duke University, Durham, NC 27708-0251, USA

Treasurer: Jiashun Jin, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

Program Secretary: Annie Qu, Department of Statistics, University of California, Irvine, Irvine, CA 92697-3425, USA

IMS PUBLICATIONS

The Annals of Statistics. *Editors*: Enno Mammen, Institute for Mathematics, Heidelberg University, 69120 Heidelberg, Germany. Lan Wang, Miami Herbert Business School, University of Miami, Coral Gables, FL 33124, USA

The Annals of Applied Statistics. *Editor-In-Chief*: Ji Zhu, Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

The Annals of Probability. *Editors*: Paul Bourgade, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012-1185, USA. Julien Dubedat, Department of Mathematics, Columbia University, New York, NY 10027, USA

The Annals of Applied Probability. *Editors*: Kavita Ramanan, Division of Applied Mathematics, Brown University, Providence, RI 02912, USA. Qi-Man Shao, Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, Guangdong 518055, P.R. China

Statistical Science. *Editor*: Moulinath Banerjee, Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

The IMS Bulletin. *Editor*: Tati Howell, bulletin@imstat.org

The Annals of Applied Statistics **[ISSN 1932-6157 (print); ISSN 1941-7330 (online)],** Volume 18, Number 1, March 2024. Published quarterly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage pending at Cleveland, Ohio, and at additional mailing offices.

POSTMASTER: Send address changes to *The Annals of Applied Statistics*, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, Maryland 21769, USA.

Copyright © 2024 by the Institute of Mathematical Statistics
Printed in the United States of America

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 1–22
https://doi.org/10.1214/23-AOAS1761

RZIMM-SCRNA: A REGULARIZED ZERO-INFLATED MIXTURE MODEL FRAMEWORK FOR SINGLE-CELL RNA-SEQ DATA

BY XINLEI MI, WILLIAM BEKERMAN, ANIL K. RUSTGI, PETER A. SIMS, PETER D. CANOLL AND JIANHUA HU[a]

Department of Biostatistics, Columbia University, [a]*jh3992@cumc.columbia.edu*

Applications of single-cell RNA sequencing in various biomedical research areas have been blooming. This new technology provides unprecedented opportunities to study disease heterogeneity at the cellular level. However, unique characteristics of scRNA-seq data, including large dimensionality, high dropout rates, and possibly batch effects, bring great difficulty into the analysis of such data. Not appropriately addressing these issues obstructs true scientific discovery. Herein we propose a unified Regularized Zero-inflated Mixture Model framework, designed for scRNA-seq data (RZiMM-scRNA), to simultaneously detect cell subgroups and identify gene differential expression based on a developed importance score, accounting for both dropouts and batch effects. We conduct extensive simulation studies in which we evaluate the performance of RZiMM-scRNA and compare it with several popular methods, including Seurat, SC3, K-means, and hierarchical clustering. Simulation results show that RZiMM-scRNA demonstrates superior clustering performance and enhanced biomarker detection accuracy, compared to alternative methods, especially when cell subgroups are less distinct, verifying the robustness of our method.

Our empirical investigations focus on two brain tumor studies dealing with astrocytoma of various grades, including the most malignant of all brain tumors, glioblastoma multiforme (GBM). Our goal is to delineate cell heterogeneity and identify driving biomarkers associated with these tumors. Notably, RZiMM-scNRA successfully identifies a small group of oligodendrocyte cells, which has drawn much attention in biomedical literature on brain cancers. In addition, our method discovers several new biomarkers which are not discussed in the original studies, including PLP1, BCAN, and PTPRZ1—all associated with the development and malignant growth of glioma—as well as CAMK2B, which is downregulated in glioma and GBM and implicated in neurodevelopment, brain function, learning and memory processes.

REFERENCES

ABELS, E. R., MAAS, S. L. N., NIELAND, L., WEI, Z., CHEAH, P. S., TAI, E., KOLSTEEG, C.-J., DUSOSWA, S. A., TING, D. A. et al. (2019). Glioblastoma-associated microglia reprogramming is mediated by functional transfer of extracellular miR-21. *Cell Rep*. **28** 3105–3119.

ARIMA-YOSHIDA, F., RAVEAU, M., SHIMOHATA, A., AMANO, K., FUKUSHIMA, A., WATANAVE, M., KOBAYASHI, S., HATTORI, S., USUI, M. et al. (2020). Impairment of spatial memory accuracy improved by Cbr1 copy number resumption and GABA B receptor-dependent enhancement of synaptic inhibition in Down syndrome model mice. *Sci*. *Rep*. **10** 1–15.

AVLIYAKULOV, N. K., RAJAVEL, K. S., LE, K. M. T., GUO, L., MIRSADRAEI, L., YONG, W. H., LIAU, L. M., LI, S., LAI, A. et al. (2014). C-terminally truncated form of αB-crystallin is associated with IDH1 R132H mutation in anaplastic astrocytoma. *J. Neuro-Oncol*. **117** 53–65.

BALCZ, B., KIRCHNER, L., CAIRNS, N., FOUNTOULAKIS, M. and LUBEC, G. (2001). Increased brain protein levels of carbonyl reductase and alcohol dehydrogenase in Down syndrome and Alzheimer's disease. In *Protein Expression in Down Syndrome Brain* 193–201. Springer, New York.

BARASH, U., SPYROU, A., LIU, P., VLODAVSKY, E., ZHU, C., LUO, J., SU, D., ILAN, N., FORSBERG-NILSSON, K. et al. (2019). Heparanase promotes glioma progression via enhancing CD24 expression. *Int. J. Cancer* **145** 1596–1608.

Key words and phrases. Single cell, mixture model, clustering, batch effect, dropout.

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M. et al. (2012). NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Res.* **41** D991–D995.

Bayer, T. A., Paliga, K., Weggen, S., Wiestler, O. D., Beyreuther, K. and Multhaup, G. (1997). Amyloid precursor-like protein 1 accumulates in neuritic plaques in Alzheimer's disease. *Acta Neuropathol.* **94** 519–524.

Butkevich, E., Härtig, W., Nikolov, M., Erck, C., Grosche, J., Urlaub, H., Schmidt, C. F., Klopfenstein, D. R. and Chua, J. J. E. (2016). Phosphorylation of FEZ1 by Microtubule Affinity Regulating Kinases regulates its function in presynaptic protein trafficking. *Sci. Rep.* **6** 1–15.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36** 411–420. https://doi.org/10.1038/nbt.4096

Büttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A., Theis, F. J. et al. (2017). Assessment of batch-correction methods for scRNA-seq data with a new test metric. BioRxiv 200345.

Byrns, C. N., Pitts, M. W., Gilman, C. A., Hashimoto, A. C. and Berry, M. J. (2014). Mice lacking selenoprotein P and selenocysteine lyase exhibit severe neurological dysfunction, neurodegeneration, and audiogenic seizures. *J. Biol. Chem.* **289** 9662–9674. https://doi.org/10.1074/jbc.M113.540682

Byun, J., Schwartz, A. G., Lusk, C., Wenzlaff, A. S., De Andrade, M., Mandal, D., Gaba, C., Yang, P., You, M. et al. (2018). Genome-wide association study of familial lung cancer. *Carcinogenesis* **39** 1135–1140.

Caballero-Franco, C. and Kissler, S. (2016). The autoimmunity-associated gene RGS1 affects the frequency of T follicular helper cells. *Genes Immun.* **17** 228–238.

Chen, J., Hou, C., Zheng, Z., Lin, H. and Lv, G. (2019a). Identification of secreted phosphoprotein 1 (SPP1) as a prognostic factor in lower-grade gliomas. *World Neurosurg.* **130** e775–e785.

Chen, Y. S., Guo, Y., Shen, H. W., Zhang, P. and Chen, H. (2019b). Screening function enrichment and related signaling pathways bioinformatics analysis of differentially expressed gene in glioma. *Zhonghua Yi Xue Za Zhi* **99** 2311–2314.

de Sousa, J. F., Torrieri, R., Serafim, R. B., Di Cristofaro, L. F. M., Escanfella, F. D., Ribeiro, R., Zanette, D. L., Paçó-Larson, M. L., da Silva Jr, W. A. et al. (2017). Expression signatures of DNA repair genes correlate with survival prognosis of astrocytoma patients. *Tumor Biol.* **39**. https://doi.org/10.1177/1010428317694552

Fraile, J. M., Campos-Iglesias, D., Rodríguez, F. and Español, Y. (2016). The deubiquitinase USP54 is overexpressed in colorectal cancer stem cells and promotes intestinal tumorigenesis. *Oncotarget* **7** 74427.

George, A. J., Gordon, L., Beissbarth, T., Koukoulas, I., Holsinger, R. M. D., Perreau, V., Cappai, R., Tan, S.-S., Masters, C. L. et al. (2010). A serial analysis of gene expression profile of the Alzheimer's disease Tg2576 mouse model. *Neurotox. Res.* **17** 360–379.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H. and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525** 251–255. https://doi.org/10.1038/nature14966

Haghverdi, L., Lun, A. T. L., Morgan, M. D. and Marioni, J. C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36** 421–427.

Hajj, G. N. M., da Silva, F. F., de Bellis, B., Lupinacci, F. C. S., Bellato, H. M., Cruz, J. R., Segundo, C. N. C., Faquini, I. V., Torres, L. C. et al. (2020). Aberrant expression of RSK1 characterizes high-grade gliomas with immune infiltration. *Mol. Oncol.* **14** 159–179.

Hedlund, E. and Deng, Q. (2018). Single-cell RNA sequencing: Technical advancements and biological applications. *Mol. Aspects Med.* **59** 36–46. https://doi.org/10.1016/j.mam.2017.07.003

Hide, T., Komohara, Y., Miyasato, Y., Nakamura, H., Makino, K., Takeya, M., Kuratsu, J.-I., Mukasa, A. and Yano, S. (2018). Oligodendrocyte progenitor cells and macrophages/microglia produce glioma stem cell niches at the tumor border. *eBioMedicine* **30** 94–104. https://doi.org/10.1016/j.ebiom.2018.02.024

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

Jia, D., Lin, W., Tang, H., Cheng, Y., Xu, K., He, Y. and Geng, W. (2019). Integrative analysis of DNA methylation and gene expression to identify key epigenetic genes in glioblastoma. *Aging* **11** 5579.

Johansson, F. K., Göransson, H. and Westermark, B. (2005). Expression analysis of genes involved in brain tumor progression driven by retroviral insertional mutagenesis in mice. *Oncogene* **24** 3896–3905. https://doi.org/10.1038/sj.onc.1208553

Johnson, W. E., Li, C. and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8** 118–127. https://doi.org/10.1093/biostatistics/kxj037

Kassambara, A. and Mundt, F. (2019). factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.6.

KHALIL, M., RIEDLBAUER, B., LANGKAMMER, C., ENZINGER, C., ROPELE, S., STOJAKOVIC, T., SCHARNAGL, H., CULEA, V., PETZOLD, A. et al. (2014). Cerebrospinal fluid transferrin levels are reduced in patients with early multiple sclerosis. *Int. MS J.* **20** 1569–1577.

KIM, J. M., SONG, J. S., CHO, H. H., SHIN, K. K., BAE, Y. C., LEE, B. J. and JUNG, J. S. (2011). Effect of the modulation of leucine zipper tumor suppressor 2 expression on proliferation of various cancer cells functions as a tumor suppressor. *Mol. Cell. Biochem.* **346** 125–136.

KISELEV, V. Y., ANDREWS, T. S. and HEMBERG, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* **20** 273–282. https://doi.org/10.1038/s41576-018-0088-9

KISELEV, V. Y., KIRSCHNER, K., SCHAUB, M. T., ANDREWS, T., YIU, A., CHANDRA, T., NATARAJAN, K. N., REIK, W., BARAHONA, M. et al. (2017). SC3: Consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14** 483–486.

KORE, R. A. and ABRAHAM, E. C. (2014). Inflammatory cytokines, interleukin-1 beta and tumor necrosis factor-alpha, upregulated in glioblastoma multiforme, raise the levels of CRYAB in exosomes secreted by U373 glioma cells. *Biochem. Biophys. Res. Commun.* **453** 326–331. https://doi.org/10.1016/j.bbrc.2014.09.068

LANGE, K. (2013). *Optimization*, 2nd ed. *Springer Texts in Statistics* **95**. Springer, New York. MR3052733 https://doi.org/10.1007/978-1-4614-5838-8

LEITE, F. A., LIRA, R. C. P., FEDATTO, P. F., ANTONINI, S. R. R., MARTINELLI JR, C. E., DE CASTRO, M., NEDER, L., RAMALHO, L. N. Z., TUCCI JR, S. et al. (2014). Low expression of HLA-DRA, HLA-DPA1, and HLA-DPB1 is associated with poor prognosis in pediatric adrenocortical tumors (ACT). *Pediatr. Blood & Cancer* **61** 1940–1948.

LI, B. and DEWEY, C. N. (2011). RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinform.* **12** 323. https://doi.org/10.1186/1471-2105-12-323

LIN, P., TROUP, M. and HO, J. W. K. (2017). CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biol.* **18**.

LINDNER, I., HEMDAN, N. Y. A., BUCHOLD, M., HUSE, K., BIGL, M., OERLECKE, I., RICKEN, A., GAUNITZ, F., SACK, U. et al. (2010). α2-macroglobulin inhibits the malignant properties of astrocytoma cells by impeding β-catenin signaling. *Cancer Res.* **70** 277–287.

MA, C., GU, C., HUO, Y., LI, X. and LUO, X.-J. (2018). The integrated landscape of causal genes and pathways in schizophrenia. *Transl. Psychiatry* **8** 1–14.

MAHERAS, K. J., PEPPI, M., GHODDOUSSI, F. and GALLOWAY, M. P. (2018). Absence of claudin 11 in CNS myelin perturbs behavior and neurotransmitter levels in mice. *Sci. Rep.* **8** 1–16.

MANGOGNA, A., BELMONTE, B., AGOSTINIS, C., ZACCHI, P., IACOPINO, D. G., MARTORANA, A., RODOLICO, V., BONAZZA, D., ZANCONATI, F. et al. (2019). Prognostic implications of the complement protein C1q in gliomas. *Front. Immunol.* **10** 2366.

METSU, S., RAINGER, J. K., DEBACKER, K. and BERNHARD, B. (2014). A CGG-repeat expansion mutation in ZNF 713 causes FRA 7 A: Association with autistic spectrum disorder in two families. *Human Mutat.* **35** 1295–1300.

MI, X., BEKERMAN, W., RUSTGI, A. K., SIMS, P. A., CANOLL, P. D. and HU, J. (2024). Supplement to “RZiMM-scRNA: A regularized zero-inflated mixture model framework for single-cell RNA-seq data.” https://doi.org/10.1214/23-AOAS1761SUPPA, https://doi.org/10.1214/23-AOAS1761SUPPB

MITTELBRONN, M., SIMON, P., LÖFFLER, C., CAPPER, D. and BUNZ, B. (2007). Elevated HLA-E levels in human glioblastomas but not in grade I to III astrocytomas correlate with infiltrating CD8+ cells. *J. Neuroimmunol.* **189** 50–58.

MIYAUCHI, E., FURUTA, T., OHTSUKI, S., TACHIKAWA, M., UCHIDA, Y., SABIT, H., OBUCHI, W., BABA, T., WATANABE, M. et al. (2018). Identification of blood biomarkers in glioblastoma by SWATH mass spectrometry and quantitative targeted absolute proteomics. *PLoS ONE* **13**.

MORAL-MORALES, D., GONZÁLEZ-OROZCO, J. C., CAPETILLO-VELÁZQUEZ, J. M., PIÑA-MEDINA, A. G., CAMACHO-ARROYO, I. et al. (2020). The role of mPRδ and mPRε in human glioblastoma cells: Expression, hormonal regulation, and possible clinical outcome. *Horm. Cancer* **11** 117–127.

MURAKAMI, A., FUKUSHIMA, C., YOSHIDOMI, K., SUEOKA, K., NAWATA, S., YOKOYAMA, Y., TSUCHIDA, S., ISMAIL, E., AL-MULLA, F. et al. (2011). Suppression of carbonyl reductase expression enhances malignant behaviour in uterine cervical squamous cell carcinoma: Carbonyl reductase predicts prognosis and lymph node metastasis. *Cancer Lett.* **311** 77–84. https://doi.org/10.1016/j.canlet.2011.06.036

NICOLE, O. and PACARY, E. (2020). CaMKIIβ in neuronal development and plasticity: An emerging candidate in brain diseases. *Int. J. Mol. Sci.* **21**. https://doi.org/10.3390/ijms21197272

O’LEARY, N. A., WRIGHT, M. W., BRISTER, J. R., CIUFO, S., HADDAD, D., MCVEIGH, R., RAJPUT, B., ROBBERTSE, B., SMITH-WHITE, B. et al. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44** D733–D745.

PETEGROSSO, R., LI, Z. and KUANG, R. (2019). Machine learning and statistical methods for clustering single-cell RNA-sequencing data. *Brief. Bioinform.*

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., Reyes, A. P., Linnarsson, S., Sandberg, R. et al. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165** 1012–1026.

Peyrot, W. J. and Price, A. L. (2021). Identifying loci with different allele frequencies among cases of eight psychiatric disorders using CC-GWAS. *Nat. Genet.* **53** 445–454. https://doi.org/10.1038/s41588-021-00787-1

Picelli, S., Faridani, O. R., Björklund, Å. K., Winberg, G. and Sagasser, S. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9** 171.

Pierson, E. and Yau, C. (2015). ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. *Genome Biol.* **16** 241. https://doi.org/10.1186/s13059-015-0805-z

Prakadan, S. M., Shalek, A. K. and Weitz, D. A. (2017). Scaling by shrinking: Empowering single-cell 'omics' with microfluidic devices. *Nat. Rev. Genet.* **18** 345–361. https://doi.org/10.1038/nrg.2017.15

R Core Team (2018). *R*: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reinius, B., Mold, J. E., Ramsköld, D., Deng, Q., Johnsson, P., Michaëlsson, J., Frisén, J. and Sandberg, R. (2016). Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nat. Genet.* **48** 1430–1435. https://doi.org/10.1038/ng.3678

Risso, D., Perraudeau, F., Gribkova, S. and Dudoit, S. (2018). A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat. Commun.* **9** 1–17.

Rousseau, A., Nutt, C. L., Betensky, R. A., Iafrate, A. J., Han, M., Ligon, K. L., Rowitch, D. H. and Louis, D. N. (2006). Expression of oligodendroglial and astrocytic lineage markers in diffuse gliomas: Use of YKL-40, ApoE, ASCL1, and NKX2-2. *J. Neuropathol. Exp. Neurol.* **65** 1149–1156.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20** 53–65.

Sawcer, S., Hellenthal, G., Pirinen, M. Spencer, C. A., Patsopoulos, N. A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C. et al. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476** 214.

Shi, Y., Ping, Y., Zhou, W., He, Z., Chen, C., Zhang, L., Chen, L., Lan, X., Zhang, X. et al. (2017). Tumour-associated macrophages secrete pleiotrophin to promote PTPRZ1 signalling in glioblastoma stem cells for tumour growth. *Nat. Commun.* **8** 1–17.

Sinmaz, N., Nguyen, T., Tea, F., Dale, R. C. and Brilot, F. (2016). Mapping autoantigen epitopes: Molecular insights into autoantibody-associated disorders of the nervous system. *J. Neuroinflamm.* **13** 219. https://doi.org/10.1186/s12974-016-0678-4

Sinreih, M., Knific, T., Thomas, P., Grazio, S. F. and Rižner, T. L. (2018). Membrane progesterone receptors β and γ have potential as prognostic biomarkers of endometrial cancer. *J. Steroid Biochem. Mol. Biol.* **178** 303–311.

Stegle, O., Teichmann, S. A. and Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16** 133–145. https://doi.org/10.1038/nrg3833

Szulzewsky, F., Pelz, A., Feng, X., Synowitz, M., Markovic, D., Langmann, T., Holtman, I. R., Wang, X., Eggen, B. J. L. et al. (2015). Glioma-associated microglia/macrophages display an expression profile different from M1 and M2 polarization and highly express Gpnmb and Spp1. *PLoS ONE* **10**.

Thyssen, G., Li, T., Lehmann, L., Zhuo, M. and Sharma, M. (2006). LZTS2 is a novel β-catenin-interacting protein and regulates the nuclear export of β-catenin. *Mol. Cell. Biol.* **26** 8857–8867.

Tibshirani, R., Walther, G. and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 411–423. MR1841503 https://doi.org/10.1111/1467-9868.00293

Treutlein, B., Brownfield, D. G., Wu, A. R., Neff, N. F., Mantalas, G. L., Espinoza, F. H., Desai, T. J., Krasnow, M. A. and Quake, S. R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509** 371–375. https://doi.org/10.1038/nature13173

Tung, P.-Y., Blischak, J. D., Hsiao, C. J., Knowles, D. A., Burnett, J. E., Pritchard, J. K. and Gilad, Y. (2017). Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep.* **7** 39921. https://doi.org/10.1038/srep39921

Vallejos, C. A., Risso, D., Scialdone, A., Dudoit, S. and Marioni, J. C. (2017). Normalizing single-cell RNA sequencing data: Challenges and opportunities. *Nat. Methods* **14** 565–571. https://doi.org/10.1038/nmeth.4292

Venteicher, A. S., Tirosh, I., Hebert, C., Yizhak, K. and Neftel, C. (2017). Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355** eaai8478.

Vidyarthi, A., Agnihotri, T., Khan, N., Singh, S., Tewari, M. K., Radotra, B. D., Chatterjee, D. and Agrewala, J. N. (2019). Predominance of M2 macrophages in gliomas leads to the suppression of local and systemic immunity. *Cancer Immunol. Immunother.* **68** 1995–2004.

WANG, D. and BODOVITZ, S. (2010). Single cell analysis: The new frontier in 'omics'. *Trends Biotechnol.* **28** 281–290. https://doi.org/10.1016/j.tibtech.2010.03.002

WANG, H., LI, B. and LENG, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 671–683. MR2749913 https://doi.org/10.1111/j.1467-9868.2008.00693.x

WANG, J., FENG, X., BAI, Z., JIN, L.-W., DUAN, Y. and LEI, H. (2014). Chromosome 19p in Alzheimer's disease: When genome meets transcriptome. *J. Alzheimer's Dis.* **38** 245–250. https://doi.org/10.3233/JAD-130917

WANG, S.-M., LIN, H.-Y., CHEN, Y.-L., HSU, T.-I., CHUANG, J.-Y., KAO, T.-J. and KO, C.-Y. (2019). CCAAT/enhancer-binding protein delta regulates the stemness of glioma stem-like cells through activating PDGFA expression upon inflammatory stimulation. *J. Neuroinflamm.* **16** 146. https://doi.org/10.1186/s12974-019-1535-z

XIA, Z., OUYANG, D., LI, Q., LI, M., ZOU, Q., LI, L. and YI, W. (2019). The expression, functions, interactions and prognostic values of PTPRZ1: A review and bioinformatic analysis. *J. Cancer* **10** 1663.

XIONG, D.-D., XU, W.-Q., HE, R.-Q., DANG, Y.-W., CHEN, G. and LUO, D.-Z. (2019). In silico analysis identified miRNA-based therapeutic agents against glioblastoma multiforme. *Oncol. Rep.* **41** 2194–2208. https://doi.org/10.3892/or.2019.7022

XU, J., CAI, L., LIAO, B. and ZHU, W. (2020). CMF-impute: An accurate imputation tool for single-cell RNA-seq data. *Bioinformatics*.

XUE, Z., HUANG, K., CAI, C., CAI, L., JIANG, C., FENG, Y., LIU, Z., ZENG, Q., CHENG, L. et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500** 593–597.

YUAN, J., LEVITIN, H. M., FRATTINI, V., BUSH, E. C., BOYETT, D. M., SAMANAMUD, J., CECCARELLI, M., DOVAS, A., ZANAZZI, G. et al. (2018). Single-cell transcriptome analysis of lineage diversity in high-grade glioma. *Gen. Med.* **10** 1–15.

ZEINER, P. S., PREUSSE, C., BLANK, A., ZACHSKORN, C., BAUMGARTEN, P., CASPARY, L., BRACZYNSKI, A. K., WEISSENBERGER, J., BRATZKE, H. et al. (2015). MIF receptor CD 74 is restricted to microglia/macrophages, associated with a M 1-polarized immune milieu and prolonged patient survival in gliomas. *Brain Pathol.* **25** 491–504.

ZEINER, P. S., PREUSSE, C., GOLEBIEWSKA, A., ZINKE, J., IRIONDO, A., MULLER, A., KAOMA, T. and FILIPSKI, K. (2019). Distribution and prognostic impact of microglia/macrophage subpopulations in gliomas. *Brain Pathol.* **29** 513–529.

ZHAN, X., JICKLING, G. C., ANDER, B. P., STAMOVA, B., LIU, D., KAO, P. F., ZELIN, M. A., JIN, L.-W., DECARLI, C. et al. (2015). Myelin basic protein associates with AßPP, Aß1-42, and amyloid plaques in cortex of Alzheimer's disease brain. *J. Alzheimer's Dis.* **44** 1213–1229.

ZHANG, D., QU, L., MA, L., ZHOU, Y., WANG, G., ZHAO, X., ZHANG, C., ZHANG, Y., WANG, M. et al. (2018). Genome-wide identification of transcription factors that are critical to non-small cell lung cancer. *Cancer Lett.* **434** 132–143.

ZHANG, F., WU, Y. and TIAN, W. (2019). A novel approach to remove the batch effect of single-cell data. *Cell Discov.* **5** 1–4.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 23–41
https://doi.org/10.1214/23-AOAS1770

SENSITIVITY ANALYSIS OF WIND ENERGY RESOURCES WITH BAYESIAN NON-GAUSSIAN AND NONSTATIONARY FUNCTIONAL ANOVA

BY JIACHEN ZHANG[1,a], PAOLA CRIPPA[2,c], MARC G. GENTON[3,d] AND STEFANO CASTRUCCIO[1,b]

[1]*Department of Applied and Computational Mathematics and Statistics, University of Notre Dame,* [a]*jzhang19@alumni.nd.edu,* [b]*scastruc@nd.edu*

[2]*Department of Civil and Environmental Engineering & Geosciences, University of Notre Dame,* [c]*pcrippa@nd.edu*

[3]*Statistics Program, King Abdullah University of Science and Technology,* [d]*marc.genton@kaust.edu.sa*

The transition from nonrenewable to renewable energies represents a global societal challenge, and developing a sustainable energy portfolio is an especially daunting task for developing countries where little to no information is available regarding the abundance of renewable resources such as wind. Weather model simulations are key to obtain such information when observational data are scarce and sparse over a country as large and geographically diverse as Saudi Arabia. However, output from such models is uncertain, as it depends on inputs such as the parametrization of the physical processes and the spatial resolution of the simulated domain. In such situations a sensitivity analysis must be performed, and the input may have a spatially heterogeneous influence of wind. In this work we propose a latent Gaussian functional analysis of variance (ANOVA) model that relies on a nonstationary Gaussian Markov random field approximation of a continuous latent process. The proposed approach is able to capture the local sensitivity of Gaussian and non-Gaussian wind characteristics such as speed and threshold exceedances over a large simulation domain, and a continuous underlying process also allows us to assess the effect of different spatial resolutions. Our results indicate that: (1) the nonlocal planetary boundary layer scheme and high spatial resolution are both instrumental in capturing wind speed and energy (especially over complex mountainous terrain), and (2) the impact of planetary boundary layer scheme and resolution on Saudi Arabia's planned wind farms is small (at most 1.4%). Thus, our results lend support for the construction of these wind farms in the next decade.

REFERENCES

AGHAHOSSEINI, A., BOGDANOV, D. and BREYER, C. (2020). Towards sustainable development in the MENA region: Analysing the feasibility of a 100% renewable electricity system in 2030. *Energy Strategy Reviews* **28** 100466.

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

BAKKA, H., VANHATALO, J., ILLIAN, J. B., SIMPSON, D. and RUE, H. (2019). Non-stationary Gaussian models with physical barriers. *Spat. Stat.* **29** 268–288. MR3903698 https://doi.org/10.1016/j.spasta.2019.01.002

BOLIN, D. and LINDGREN, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *Ann. Appl. Stat.* **5** 523–550. MR2810408 https://doi.org/10.1214/10-AOAS383

CASTRUCCIO, S. and GUINNESS, J. (2017). An evolutionary spectrum approach to incorporate large-scale geographical descriptors on global processes. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 329–344. MR3611690 https://doi.org/10.1111/rssc.12167

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

Key words and phrases. Bayesian hierarchical model, stochastic partial differential equation, functional ANOVA, wind energy.

CRESSIE, N. and HUANG, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.* **94** 1330–1340. MR1731494 https://doi.org/10.2307/2669946

CRIPPA, P., ALIFA, M., BOLSTER, D., GENTON, M. G. and CASTRUCCIO, S. (2021). A temporal model for vertical extrapolation of wind speed and wind energy assessment. *Appl. Energy* **301** 117378.

DINCER, I. (2018). 1.12 fossil fuels. In *Comprehensive Energy Systems* (I. Dincer, ed.) 521–567. Elsevier, Oxford.

EDWARDS, M., CASTRUCCIO, S. and HAMMERLING, D. (2020). Marginally parameterized spatio-temporal models and stepwise maximum likelihood estimation. *Comput. Statist. Data Anal.* **151** 107018, 12. MR4107728 https://doi.org/10.1016/j.csda.2020.107018

EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS (2016). ECMWF IFS CY41r2 high-resolution operational forecasts. https://doi.org/10.5065/D68050ZV

FUGLSTAD, G.-A. and CASTRUCCIO, S. (2020). Compression of climate simulations with a nonstationary global spatiotemporal SPDE model. *Ann. Appl. Stat.* **14** 542–559. MR4117819 https://doi.org/10.1214/20-AOAS1340

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

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

GIANI, P., TAGLE, F., GENTON, M. G., CASTRUCCIO, S. and CRIPPA, P. (2020). Closing the gap between wind energy targets and implementation for emerging countries. *Appl. Energy* **269** 115085.

GUALTIERI, G. (2019). A comprehensive review on wind resource extrapolation models applied in wind energy. *Renew. Sustain. Energy Rev.* **102** 215–233.

HASANEAN, H. and ALMAZROUI, M. (2015). Rainfall: Features and variations over Saudi Arabia, a review. *Climate* **3** 578–626.

HU, W., FUGLSTAD, G.-A. and CASTRUCCIO, S. (2022). A stochastic locally diffusive model with neural network-based deformations for global sea surface temperature. *Stat* **11** Paper No. e431, 9. MR4394991 https://doi.org/10.1002/sta4.431

HU, X., SIMPSON, D., LINDGREN, F. and RUE, H. (2013). Multivariate Gaussian random fields using systems of stochastic partial differential equations. Available at arXiv:1307.1379.

HU, X.-M., NIELSEN-GAMMON, J. W. and ZHANG, F. (2010). Evaluation of three planetary boundary layer schemes in the WRF model. *J. Appl. Meteorol. Climatol.* **49** 1831–1844.

HUANG, J. Z., KOOPERBERG, C., STONE, C. J. and TRUONG, Y. K. (2000). Functional ANOVA modeling for proportional hazards regression. *Ann. Statist.* **28** 961–999. MR1810916 https://doi.org/10.1214/aos/1015956703

JANJIĆ, Z. I. (1994). The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Weather Rev.* **122** 927–945.

JEONG, J., CASTRUCCIO, S., CRIPPA, P. and GENTON, M. G. (2018). Reducing storage of global wind ensembles with stochastic generators. *Ann. Appl. Stat.* **12** 490–509. MR3773402 https://doi.org/10.1214/17-AOAS1105

JEONG, J., YAN, Y., CASTRUCCIO, S. and GENTON, M. G. (2019). A stochastic generator of global monthly wind energy with Tukey g-and-h autoregressive processes. *Statist. Sinica* **29** 1105–1126. MR3932511

JONES, C., GIORGI, F. and ASRAR, G. (2011). The coordinated regional downscaling experiment: Cordex–an international downscaling link to cmip5. *CLIVAR Exchanges* **16** 34–40.

JUN, M. and STEIN, M. L. (2008). Nonstationary covariance models for global data. *Ann. Appl. Stat.* **2** 1271–1289. MR2655659 https://doi.org/10.1214/08-AOAS183

KAUFMAN, C. G. and SAIN, S. R. (2010). Bayesian functional ANOVA modeling using Gaussian process prior distributions. *Bayesian Anal.* **5** 123–149. MR2596438 https://doi.org/10.1214/10-BA505

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

MOHSENI-CHERAGHLOU, A. (2021). Fossil fuel subsidies and renewable energies in mena: An oxymoron? https://www.mei.edu/publications/fossil-fuel-subsidies-and-renewable-energies-mena-oxymoron.

NREP (2018). Saudi arabia renewable energy targets and long term visibility. national renewable energy program.

NURUNNABI, M. (2017). Transformation from an oil-based economy to a knowledge-based economy in Saudi Arabia: The direction of saudi vision 2030. *J. Knowl. Econ.* **8** 536–64.

PETERSON, E. W. and HENNESSEY, J. P. (1978). On the use of power laws for estimates of wind power potential. *J. Appl. Meteorol.* **17** 390–394.

PLEIM, J. E. (2007). A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *J. Appl. Meteorol. Climatol.* **46** 1383–1395.

QU, Z., DAI, W. and GENTON, M. G. (2021). Robust functional multivariate analysis of variance with environmental applications. *Environmetrics* **32** Paper No. e2641, 23. MR4207556 https://doi.org/10.1002/env.2641

REN21 SECRETARIAT (2021). Renewables 2021 – global status report. Paris, France.

RENEWABLES, R. (2018). Global status report (Paris: Ren21 secretariat).

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

SAIN, S. R., NYCHKA, D. and MEARNS, L. (2011). Functional ANOVA and regional climate experiments: A statistical analysis of dynamic downscaling. *Environmetrics* **22** 700–711. MR2843137 https://doi.org/10.1002/env.1068

SIUTA, D., WEST, G. and STULL, R. (2017). WRF hub-height wind forecast sensitivity to pbl scheme, grid length, and initial condition choice in complex terrain. *Weather Forecast.* **32** 493–509.

SKAMAROCK, W. C., KLEMP, J. B., DUDHIA, J., GILL, D. O., LIU, Z., BERNER, J., WANG, W., POWERS, J. G., DUDA, M. G. et al. (2019). *A Description of the Advanced Research Wrf Model Version* 4. National Center for Atmospheric Research, Boulder, CO, USA.

STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. *Springer Series in Statistics*. Springer, New York. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

STONE, C. J., HANSEN, M. H., KOOPERBERG, C. and TRUONG, Y. K. (1997). Polynomial splines and their tensor products in extended linear modeling. *Ann. Statist.* **25** 1371–1470. MR1463561 https://doi.org/10.1214/aos/1031594728

SUN, Y. and GENTON, M. G. (2012). Functional median Polish. *J. Agric. Biol. Environ. Stat.* **17** 354–376. MR2993271 https://doi.org/10.1007/s13253-012-0096-8

TAGLE, F., CASTRUCCIO, S., CRIPPA, P. and GENTON, M. G. (2019). A non-Gaussian spatio-temporal model for daily wind speeds based on a multi-variate skew-t distribution. *J. Time Series Anal.* **40** 312–326. MR3946155 https://doi.org/10.1111/jtsa.12437

TARRÍO-SAAVEDRA, J., NAYA, S., FRANCISCO-FERNÁNDEZ, M., ARTIAGA, R. and LOPEZ-BECEIRO, J. (2011). Application of functional anova to the study of thermal stability of micro–nano silica epoxy composites. *Chemom. Intell. Lab. Syst.* **105** 114–124.

ULLAH, S. and FINCH, C. (2013). Applications of functional data analysis: A systematic review. *BMC Med. Res. Methodol.* **13**.

WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41** 434–449. MR0067450 https://doi.org/10.1093/biomet/41.3-4.434

WHITTLE, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika* **50** 129–134. MR0161430 https://doi.org/10.1093/biomet/50.1-2.129

YUE, Y., BOLIN, D., RUE, H. and WANG, X.-F. (2019). Bayesian generalized two-way ANOVA modeling for functional data using INLA. *Statist. Sinica* **29** 741–767. MR3931386

YUE, Y. and SPECKMAN, P. L. (2010). Nonstationary spatial Gaussian Markov random fields. *J. Comput. Graph. Statist.* **19** 96–116. MR2654402 https://doi.org/10.1198/jcgs.2009.08124

ZHANG, J., CRIPPA, P., GENTON, M. G. and CASTRUCCIO, S. (2021). Assessing the reliability of wind power operations under a changing climate with a non-Gaussian bias correction. *Ann. Appl. Stat.* **15** 1831–1849. MR4355078 https://doi.org/10.1214/21-aoas1460

ZHANG, J.-T., CHENG, M.-Y., WU, H.-T. and ZHOU, B. (2019). A new test for functional one-way ANOVA with applications to ischemic heart screening. *Comput. Statist. Data Anal.* **132** 3–17. MR3913131 https://doi.org/10.1016/j.csda.2018.05.004

ZHANG, Y., HODGES, J. S. and BANERJEE, S. (2009). Smoothed ANOVA with spatial effects as a competitor to MCAR in multivariate spatial smoothing. *Ann. Appl. Stat.* **3** 1805–1830. MR2752159 https://doi.org/10.1214/09-AOAS267

ZHANG, J., CRIPPA, P., GENTON, M. G and CASTRUCCIO, S. (2024). Supplement to “Sensitivity analysis of wind energy resources with Bayesian non-Gaussian and nonstationary functional ANOVA.” https://doi.org/10.1214/23-AOAS1770SUPP

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 42–62
https://doi.org/10.1214/23-AOAS1778

ON THE FAIR COMPARISON OF OPTIMIZATION ALGORITHMS IN DIFFERENT MACHINES

BY ETOR ARZA[1,a], JOSU CEBERIO[2,c], EKHIÑE IRUROZKI[3,d] AND ARITZ PÉREZ[1,b]

[1]*Data Science Department, BCAM—Basque Center for Applied Mathematics,* [a]*earza@bcamath.org,* [b]*aperez@bcamath.org*
[2]*Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU,* [c]*josu.ceberio@ehu.eus*
[3]*Signal, Statistics and Learning Department, Télécom Paris,* [d]*irurozki@telecom-paris.fr*

An experimental comparison of two or more optimization algorithms requires the same computational resources to be assigned to each algorithm. When a maximum runtime is set as the stopping criterion, all algorithms need to be executed in the same machine if they are to use the same resources. Unfortunately, the implementation code of the algorithms is not always available, which means that running the algorithms to be compared in the same machine is not always possible. And even if they are available, some optimization algorithms might be costly to run, such as training large neural-networks in the cloud.

In this paper we consider the following problem: how do we compare the performance of a new optimization algorithm B with a known algorithm A in the literature if we only have the results (the objective values) and the runtime in each instance of algorithm A? Particularly, we present a methodology that enables a statistical analysis of the performance of algorithms executed in different machines. The proposed methodology has two parts. First, we propose a model that, given the runtime of an algorithm in a machine, estimates the runtime of the same algorithm in another machine. This model can be adjusted so that the probability of estimating a runtime longer than what it should be is arbitrarily low. Second, we introduce an adaptation of the one-sided sign test that uses a modified p-value and takes into account that probability. Such adaptation avoids increasing the probability of type I error associated with executing algorithms A and B in different machines.

REFERENCES

ARZA, E., CEBERIO, J., IRUROZKI, E. and PÉREZ, A. (2024). Supplement to "On the fair comparison of optimization algorithms in different machines." https://doi.org/10.1214/23-AOAS1778SUPPA, https://doi.org/10.1214/23-AOAS1778SUPPB

ARZA, E., PÉREZ, A., IRUROZKI, E. and CEBERIO, J. (2020). Kernels of Mallows models under the Hamming distance for solving the quadratic assignment problem. *Swarm Evol. Comput.* 100740. https://doi.org/10.1016/j.swevo.2020.100740

BENLIC, U. and HAO, J.-K. (2015). Memetic search for the quadratic assignment problem. *Expert Syst. Appl.* **42** 584–595. https://doi.org/10.1016/j.eswa.2014.08.011

BROWN, T. B., MANN, B., RYDER, N., SUBBIAH, M., KAPLAN, J., DHARIWAL, P., NEELAKANTAN, A., SHYAM, P., SASTRY, G., ASKELL, A., AGARWAL, S., HERBERT-VOSS, A., KRUEGER, G., HENIGHAN, T., CHILD, R., RAMESH, A., ZIEGLER, D. A., WU, J., WINTER, C., HESSE, C., CHEN, M., SIGLER, E., LITWIN, M., GRAY, S., CHESS, B., CLARK, J., BERNER, C., MCCANDLISH, S., RADFORD, A., SUTSKEVER, I. and AMODEI, D. (2020). Language models are few-shot learners. Available at arXiv:2005.14165. https://doi.org/10.48550/ARXIV.2005.14165

CONOVER, W. J. (1980). *Practical Nonparametric Statistics*.

DOMÍNGUEZ, J. and ALBA, E. (2012). A methodology for comparing the execution time of metaheuristics running on different hardware. In *Evolutionary Computation in Combinatorial Optimization* **7245** 1–12. Springer, Berlin. https://doi.org/10.1007/978-3-642-29124-1_1

HUTSON, M. (2017). Artificial intelligence, in so many words. *Science* **356** 19. MR3700131 https://doi.org/10.1126/science.357.6346.19

Key words and phrases. Algorithms, optimization, benchmarking, statistical tests.

NORDHAUS, W. D. (2007). Two centuries of productivity growth in computing. *J. Econ. Hist.* **67** 128–159. https://doi.org/10.1017/S0022050707000058

SHARIR, O., PELEG, B. and SHOHAM, Y. (2020). The cost of training NLP models: A concise overview. arXiv preprint. Available at arXiv:2004.08900.

WASSERSTEIN, R. L. and LAZAR, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose [Editorial]. *Amer. Statist.* **70** 129–133. MR3511040 https://doi.org/10.1080/00031305.2016.1154108

WEICKER, R. P. (1988). Dhrystone benchmark: Rationale for version 2 and measurement rules. *ACM SIGPLAN Not.* **23** 49–62. https://doi.org/10.1145/47907.47911

ZEGGEL, M. (2019). Magic-square. Available at https://github.com/Gizmoscope/magic-square.

GEEKSFORGEEKS (2022a) 0-1 knapsack problem. https://www.geeksforgeeks.org/0-1-knapsack-problem-dp-10/.

GEEKSFORGEEKS (2022b). N queen problem. https://www.geeksforgeeks.org/n-queen-problem-backtracking-3/.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 63–79
https://doi.org/10.1214/23-AOAS1779

DISTRIBUTED PROPORTIONAL LIKELIHOOD RATIO MODEL WITH APPLICATION TO DATA INTEGRATION ACROSS CLINICAL SITES

BY CHONGLIANG LUO[1,a], RUI DUAN[2,b], MACKENZIE EDMONDSON[3,c], JIASHENG SHI[4,g], MITCHELL MALTENFORT[4,h], JEFFREY S. MORRIS[3,d], CHRISTOPHER B. FORREST[4,i], REBECCA HUBBARD[3,e] AND YONG CHEN[3,f]

[1]*Division of Public Health Sciences, Washington University School of Medicine in St. Louis,* [a]*chongliang@wustl.edu*

[2]*Department of Biostatistics, Harvard T.H. School of Public Health,* [b]*rduan@hsph.harvard.edu*

[3]*Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania,* [c]*macjohn@pennmedicine.upenn.edu,* [d]*jeffrey.morris@pennmedicine.upenn.edu,* [e]*rhubb@pennmedicine.upenn.edu,* [f]*ychen123@upenn.edu*

[4]*Applied Clinical Research Center, Roberts Center for Pediatric Research, Children's Hospital of Philadelphia,* [g]*jiasheng.shi@pennmedicine.upenn.edu,* [h]*maltenform@chop.edu,* [i]*forrestc@chop.edu*

Real-world evidence synthesis through integration of data from distributed research networks has gained increasing attention in recent years. Due to privacy concerns and restrictions of sharing patient-level data, distributed algorithms that do not require sharing patient level information are in great need for facilitating multisite collaborations. On the other hand, data collected at multiple sites often come from diverse populations, and there exists a substantial amount of heterogeneity across sites in patient characteristics. Most of the existing distributed algorithms have ignored such between-site heterogeneity. In this paper we aim to fill this methodological gap by proposing a general distributed algorithm. We develop our distributed algorithm based on a general semiparametric model, namely, the proportional likelihood ratio model (*Biometrika* **99** (2012) 211–222), which is a semiparametric extension of generalized linear model. We devise the proportional likelihood ratio model with site-specific baseline function, to account for between-site heterogeneity, and shared regression parameters to borrow information across sites. Under this flexible formulation, our distributed algorithm is designed to be privacy-preserving and communication-efficient (i.e., only one round of communication across sites is needed). We validate our method via simulation studies and demonstrate the utility of our method via a multisite study of pediatric avoidable hospitalization based on electronic health record data from a total of 354,672 patients across 26 different clinical sites within the Children's Hospital of Philadelphia health system.

REFERENCES

BUI, A. L., DIELEMAN, J. L., HAMAVID, H., BIRGER, M., CHAPIN, A., DUBER, H. C., HORST, C., REYNOLDS, A., SQUIRES, E. et al. (2017). Spending on children's personal health care in the United States, 1996-2013. *JAMA Pediatr.* **171** 181–189.

CHAN, K. C. G. (2013). Nuisance parameter elimination for proportional likelihood ratio models with nonignorable missingness and random truncation. *Biometrika* **100** 269–276. MR3034342 https://doi.org/10.1093/biomet/ass056

CHEN, Y., NING, J. and CAI, C. (2015). Regression analysis of longitudinal data with irregular and informative observation times. *Biostatistics* **16** 727–739. MR3449839 https://doi.org/10.1093/biostatistics/kxv008

COX, D. R. and REID, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91** 729–737. MR2090633 https://doi.org/10.1093/biomet/91.3.729

DU, J., PARK, Y.-T., THEERA-AMPORNPUNT, N., MCCULLOUGH, J. S. and SPEEDIE, S. M. (2012). The use of count data models in biomedical informatics evaluation research. *J. Amer. Med. Inform. Assoc.* **19** 39–44.

DUAN, R., BOLAND, M. R., LIU, Z., LIU, Y., CHANG, H. H., XU, H., CHU, H., SCHMID, C. H., FORREST, C. B. et al. (2020a). Learning from electronic health records across multiple sites: A communication-efficient and privacy-preserving distributed algorithm. *J. Amer. Med. Inform. Assoc.* **27** 376–385.

Key words and phrases. Distributed research network, heterogeneity-aware distributed algorithms, noniterative distributed algorithm, privacy-preserving, real-world evidence.

DUAN, R., LUO, C., SCHUEMIE, M. H., TONG, J., LIANG, J. C., CHANG, H. H., BOLAND, M. R., BIAN, J., XU, H. et al. (2020b). Learning from local to global: An efficient distributed algorithm for modeling time-to-event data. *J. Amer. Med. Inform. Assoc.* **27** 1028–1036.

DUAN, R., NING, Y. and CHEN, Y. (2022). Heterogeneity-aware and communication-efficient distributed statistical inference. *Biometrika* **109** 67–83. MR4374641 https://doi.org/10.1093/biomet/asab007

DUKE, J. D., RYAN, P. B., SUCHARD, M. A., HRIPCSAK, G., JIN, P., REICH, C., SCHWALM, M.-S., KHOMA, Y., WU, Y. et al. (2017). Risk of angioedema associated with levetiracetam compared with phenytoin: Findings of the observational health data sciences and informatics research network. *Epilepsia* **58** e101–e106.

EDDELBUETTEL, D. (2013). *Seamless R and C++ Integration with Rcpp*. Springer, Berlin.

EDMONDSON, M. J., LUO, C., DUAN, R., MALTENFORT, M., CHEN, Z., LOCKE JR, K., SHULTS, J., BIAN, J., RYAN, P. B. et al. (2021). An efficient and accurate distributed learning algorithm for modeling multi-site zero-inflated count outcomes. *Sci. Rep.* **11** 19647.

EDMONDSON, M. J., LUO, C., ISLAM, M. N., SHEILS, N. E., BURESH, J., CHEN, Z., BIAN, J. and CHEN, Y. (2022). Distributed quasi-Poisson regression algorithm for modeling multi-site count outcomes in distributed data networks. *J. Biomed. Inform.* **131** 104097.

FAY, M. P. and GRAUBARD, B. I. (2001). Small-sample adjustments for Wald-type tests using sandwich estimators. *Biometrics* **57** 1198–1206. MR1950428 https://doi.org/10.1111/j.0006-341X.2001.01198.x

FDA (2018). Use of electronic health record data in clinical investigations. Available at https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances.

FIRTH, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80** 27–38. MR1225212 https://doi.org/10.1093/biomet/80.1.27

FORREST, C. B., MARGOLIS, P. A., BAILEY, L. C., MARSOLO, K., DEL BECCARO, M. A., FINKELSTEIN, J. A., MILOV, D. E., VIELAND, V. J., WOLF, B. A. et al. (2014). PEDSnet: A national pediatric learning health system. *J. Amer. Med. Inform. Assoc.* **21** 602–606.

FRIEDMAN, C. P., WONG, A. K. and BLUMENTHAL, D. (2010). Achieving a nationwide learning health system. *Sci. Transl. Med.* **2** 57–29.

HRIPCSAK, G., RYAN, P. B., DUKE, J. D., SHAH, N. H., PARK, R. W., HUSER, V., SUCHARD, M. A., SCHUEMIE, M. J., DEFALCO, F. J. et al. (2016). Characterizing treatment pathways at scale using the OHDSI network. *Proc. Natl. Acad. Sci. USA* **113** 7329–7336.

JORDAN, M. I., LEE, J. D. and YANG, Y. (2019). Communication-efficient distributed statistical inference. *J. Amer. Statist. Assoc.* **114** 668–681. MR3963171 https://doi.org/10.1080/01621459.2018.1429274

KALBFLEISCH, J. D. (1978). Likelihood methods and nonparametric tests. *J. Amer. Statist. Assoc.* **73** 167–170. MR0518600

LI, P. and REDDEN, D. T. (2015). Small sample performance of bias-corrected sandwich estimators for cluster-randomized trials with binary outcomes. *Stat. Med.* **34** 281–296. MR3293148 https://doi.org/10.1002/sim.6344

LIANG, K.-Y. and QIN, J. (2000). Regression analysis under non-standard situations: A pairwise pseudolikelihood approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 773–786. MR1796291 https://doi.org/10.1111/1467-9868.00263

LIANG, K.-Y. and ZEGER, S. L. (1995). Inference based on estimating functions in the presence of nuisance parameters. *Statist. Sci.* **10** 158–173. MR1368098

LINDSAY, B. G. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes* (*Ithaca*, *NY*, 1987). *Contemp. Math.* **80** 221–239. Amer. Math. Soc., Providence, RI. MR0999014 https://doi.org/10.1090/conm/080/999014

LINES, L. M., ROSEN, A. B. and ASH, A. S. (2017). Enhancing administrative data to predict emergency department utilization: The role of neighborhood sociodemographics. *J. Health Care Poor Underserv.* **28** 1487–1508. https://doi.org/10.1353/hpu.2017.0129

LIU, X., DUAN, R., LUO, C., OGDIE, A., MOORE, J. H., KRANZLER, H. R., BIAN, J. and CHEN, Y. (2022). Multisite learning of high-dimensional heterogeneous data with applications to opioid use disorder study of 15,000 patients across 5 clinical sites. *Sci. Rep.* **12** 11073.

LU, C.-L., WANG, S., JI, Z., WU, Y., XIONG, L., JIANG, X. and OHNO-MACHADO, L. (2015). WebDISCO: A web service for distributed Cox model learning without patient-level data sharing. *J. Amer. Med. Inform. Assoc.* **22** 1212–1219.

LU, S. and KUO, D. Z. (2012). Hospital charges of potentially preventable pediatric hospitalizations. *Acad. Pediatr.* **12** 436–444. https://doi.org/10.1016/j.acap.2012.06.006

LUO, C., DUAN, R., EDMONDSON, M., SHI, J., MALTENFORT, M., MORRIS, J. S., FORREST, C. B., HUBBARD, R. and CHEN, Y. (2024). Supplement to "Distributed proportional likelihood ratio model with application to data integration across clinical sites." https://doi.org/10.1214/23-AOAS1779SUPPA, https://doi.org/10.1214/23-AOAS1779SUPPB, https://doi.org/10.1214/23-AOAS1779SUPPC, https://doi.org/10.1214/23-AOAS1779SUPPD

LUO, C., DUAN, R., NAJ, A. C., KRANZLER, H. R., BIAN, J. and CHEN, Y. (2022a). ODACH: A one-shot distributed algorithm for Cox model with heterogeneous multi-center data. *Sci. Rep.* **12** 6627.

LUO, C., ISLAM, M. N., SHEILS, N. E., BURESH, J., REPS, J., SCHUEMIE, M. J., RYAN, P. B., EDMONDSON, M., DUAN, R. et al. (2022b). DLMM as a lossless one-shot algorithm for collaborative multi-site distributed linear mixed models. *Nat. Commun.* **13** 1678.

LUO, C., ISLAM, M. N., SHEILS, N. E., BURESH, J., SCHUEMIE, M. J., DOSHI, J. A., WERNER, R. M., ASCH, D. A. and CHEN, Y. (2022c). dPQL: A lossless distributed algorithm for generalized linear mixed model with application to privacy-preserving hospital profiling. *J. Amer. Med. Inform. Assoc.* **29** 1366–1371.

LUO, X. and TSAI, W. Y. (2012). A proportional likelihood ratio model. *Biometrika* **99** 211–222. MR2899674 https://doi.org/10.1093/biomet/asr060

LUO, X. and TSAI, W. Y. (2015). Moment-type estimators for the proportional likelihood ratio model with longitudinal data. *Biometrika* **102** 121–134. MR3335100 https://doi.org/10.1093/biomet/asu055

MALTENFORT, M. G., CHEN, Y. and FORREST, C. B. (2019). Prediction of 30-day pediatric unplanned hospitalizations using the johns Hopkins adjusted clinical groups risk adjustment system. *PLoS ONE* **14** e0221233.

NEELON, B., CHANG, H. H., LING, Q. and HASTINGS, N. S. (2016). Spatiotemporal hurdle models for zero-inflated count data: Exploring trends in emergency department visits. *Stat. Methods Med. Res.* **25** 2558–2576. MR3572870 https://doi.org/10.1177/0962280214527079

NEYMAN, J. and SCOTT, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16** 1–32. MR0025113 https://doi.org/10.2307/1914288

SEN, P. K. (1960). On some convergence properties of U-statistics. *Calcutta Statist. Assoc. Bull.* **10** 1–18. MR0119286 https://doi.org/10.1177/0008068319600101

SHADDOX, T. R., RYAN, P. B., SCHUEMIE, M. J., MADIGAN, D. and SUCHARD, M. A. (2016). Hierarchical models for multiple, rare outcomes using massive observational healthcare databases. *Stat. Anal. Data Min.* **9** 260–268. MR3529398 https://doi.org/10.1002/sam.11324

SHEN, W., LIU, S., CHEN, Y. and NING, J. (2019). Regression analysis of longitudinal data with outcome-dependent sampling and informative censoring. *Scand. J. Stat.* **46** 831–847. MR3994171 https://doi.org/10.1111/sjos.12373

SHERMAN, R. E., ANDERSON, S. A., DAL PAN, G. J., GRAY, G. W., GROSS, T., HUNTER, N. L., LAVANGE, L., MARINAC-DABIC, D., MARKS, P. W. et al. (2016). Real-world evidence—what is it and what can it tell us. *N. Engl. J. Med.* **375** 2293–2297.

SHU, D., YOSHIDA, K., FIREMAN, B. H. and TOH, S. (2020). Inverse probability weighted Cox model in multi-site studies without sharing individual-level data. *Stat. Methods Med. Res.* **29** 1668–1681. MR4106964 https://doi.org/10.1177/0962280219869742

SUCHARD, M. A., SCHUEMIE, M. J., KRUMHOLZ, H. M., YOU, S. C., CHEN, R., PRATT, N., REICH, C. G., DUKE, J., MADIGAN, D. et al. (2019). Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: A systematic, multinational, large-scale analysis. *Lancet* **394** 1816–1826.

TONG, J., LUO, C., ISLAM, M. N., SHEILS, N. E., BURESH, J., EDMONDSON, M., MERKEL, P. A., LAUTENBACH, E., DUAN, R. et al. (2022). Distributed learning for heterogeneous clinical data with application to integrating Covid-19 data across 230 sites. *NPJ Dig. Med.* **5** 76.

VARIN, C., REID, N. and FIRTH, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21** 5–42. MR2796852

VASHISHT, R., JUNG, K., SCHULER, A., BANDA, J. M., PARK, R. W., JIN, S., LI, L., DUDLEY, J. T., JOHNSON, K. W. et al. (2018). Association of hemoglobin A1c levels with use of sulfonylureas, dipeptidyl peptidase 4 inhibitors, and thiazolidinediones in patients with type 2 diabetes treated with metformin: Analysis from the observational health data sciences and informatics initiative. *JAMA Netw. Open* **1** e181755–e181755.

WU, Y., JIANG, X., KIM, J. and OHNO-MACHADO, L. (2012). G rid binary LO gistic RE gression (GLORE): Building shared models without sharing data. *J. Amer. Med. Inform. Assoc.* **19** 758–764.

ZEILEIS, A., KLEIBER, C. and JACKMAN, S. (2008). Regression models for count data in R. *J. Stat. Softw.* **27** 1–25.

ZHU, H. (2014). Likelihood approaches for proportional likelihood ratio model with right-censored data. *Stat. Med.* **33** 2467–2479. MR3256678 https://doi.org/10.1002/sim.6105

ZHU, R., JIANG, C., WANG, X., WANG, S., ZHENG, H. and TANG, H. (2020). Privacy-preserving construction of generalized linear mixed model for biomedical computation. *Bioinformatics* **36** i128–i135.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 80–99
https://doi.org/10.1214/23-AOAS1780

BAYESIAN MULTIPLE INSTANCE CLASSIFICATION BASED ON HIERARCHICAL PROBIT REGRESSION

BY DANYI XIONG[1,a], SEONGOH PARK[2,b], JOHAN LIM[3,c], TAO WANG[4,d] AND XINLEI WANG[5,e]

[1]*Department of Statistical Science, Southern Methodist University,* [a]*dxiong@smu.edu*
[2]*Department of Statistics, Sungshin Women's University,* [b]*spark6@sungshin.ac.kr*
[3]*Department of Statistics, Seoul National University,* [c]*johanlim@stats.snu.ac.kr*
[4]*Quantitative Biomedical Research Center, Peter O'Donnell Jr. School of Public Health, UT Southwestern Medical Center,* [d]*Tao.Wang@utsouthwestern.edu*
[5]*Department of Mathematics, University of Texas at Arlington,* [e]*xinlei.wang@uta.edu*

In multiple instance learning (MIL), the response variable is predicted by features (or covariates) of one or more instances, which are collectively denoted as a bag. Learning the relationship between bags and instances is challenging because of the unknown and possibly complicated data generating mechanism regarding how instances contribute to the bag label. MIL has been applied to solve a variety of real-world problems, which have been mostly focused on supervised tasks, such as molecule activity prediction, protein binding affinities prediction, object detection, and computer-aided diagnosis. However, to date, the majority of the off-the-shelf MIL methods are developed in the computer science domain, and they focus on improving the prediction performance while spending little effort on explainability of the algorithm. In this article a Bayesian multiple instance learning model, based on probit regression (MICProB), is proposed, which contributes a significant portion to the suite of statistical methodologies for MIL. MICProB is composed of two nested probit regression models, where the inner model is estimated for predicting primary instances, which are considered as the "important" ones that determine the bag label, and the outer model is for predicting bag-level responses based on the primary instances estimated by the inner model. The posterior distribution of MICProB can be conveniently approximated using a Gibbs sampler, and the prediction for new bags can be performed in a fully integrated Bayesian way. We evaluate the performance of MICProB against 15 benchmark methods and demonstrate its competitiveness in simulation and real-data examples. In addition to its capability of identifying primary instances, as compared to existing optimization-based approaches, MICProB also enjoys great advantages in providing a transparent model structure, straightforward statistical inference of quantities related to model parameters, and favorable interpretability of covariate effects on the bag-level response.

REFERENCES

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679.

AMORES, J. (2013). Multiple instance classification: Review, taxonomy and comparative study. *Artificial Intelligence* **201** 81–105.

ANDREWS, S., TSOCHANTARIDIS, I. and HOFMANN, T. (2003). Support vector machines for multiple-instance learning. In *Advances in Neural Information Processing Systems* 577–584.

BABENKO, B., DOLLÁR, P., TU, Z. and BELONGIE, S. (2008). Simultaneous learning and alignment: Multi-instance and multi-pose learning. In *Workshop on Faces in 'Real-Life' Images*: *Detection*, *Alignment*, *and Recognition*.

Key words and phrases. Binary classification, Bayesian inference, Gibbs sampling, primary instance, weakly supervised learning.

BOLOTIN, D. A., SHUGAY, M., MAMEDOV, I. Z., PUTINTSEVA, E. V., TURCHANINOVA, M. A., ZVYAGIN, I. V., BRITANOVA, O. V. and CHUDAKOV, D. M. (2013). MiTCR: Software for T-cell receptor sequencing data analysis. *Nat. Methods* **10** 813–814. https://doi.org/10.1038/nmeth.2555

BYERS, L. A. and RUDIN, C. M. (2015). Small cell lung cancer: Where do we go from here? *Cancer* **121** 664–672.

CARBONNEAU, M.-A., CHEPLYGINA, V., GRANGER, E. and GAGNON, G. (2018). Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognit.* **77** 329–353.

CAWLEY, G. C. and TALBOT, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **11** 2079–2107.

CHEN, P.-Y., CHEN, C.-C., YANG, C.-H., CHANG, S.-M. and LEE, K.-J. (2017). milr: Multiple-instance logistic regression with lasso penalty. *R J.* **9** 446.

CHEN, Y., BI, J. and WANG, J. Z. (2006). MILES: Multiple-instance learning via embedded instance selection. *IEEE Trans. Pattern Anal. Mach. Intell.* **28** 1931–1947.

CHEPLYGINA, V., TAX, D. M. and LOOG, M. (2015). Multiple instance learning with bag dissimilarities. *Pattern Recognit.* **48** 264–275.

CLARKE-PEARSON, D. L. (2009). Clinical practice. Screening for ovarian cancer. *N. Engl. J. Med.* **361** 170–177. https://doi.org/10.1056/NEJMcp0901926

CORNABY, C., GIBBONS, L., MAYHEW, V., SLOAN, C. S., WELLING, A. and POOLE, B. D. (2015). B cell epitope spreading: Mechanisms and contribution to autoimmune diseases. *Immunol. Lett.* **163** 56–68. https://doi.org/10.1016/j.imlet.2014.11.001

DIETTERICH, T. G., LATHROP, R. H. and LOZANO-PÉREZ, T. (1997). Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence* **89** 31–71.

FOULDS, J. and FRANK, E. (2010). A review of multi-instance learning assumptions. *Knowl. Eng. Rev.* **25** 1–25.

GÄRTNER, T., FLACH, P. A., KOWALCZYK, A. and SMOLA, A. J. (2002). Multi-instance kernels. In *Proceedings of the* 19*th International Conference on Machine Learning* 179–186.

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2013). *Bayesian Data Analysis*. Chapman & Hall, London.

GELMAN, A., JAKULIN, A., PITTAU, M. G. and SU, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2** 1360–1383.

GELMAN, A., SU, Y.-S., YAJIMA, M., HILL, J., PITTAU, M. G., KERMAN, J., ZHENG, T., DORIE, V. and SU, M. Y.-S. (2013). Package'arm'. *Data Analysis Using Regression and Multilevel/Hierarchical Models*.

HE, H. and GARCIA, E. A. (2009). Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **21** 1263–1284.

HELD, L. and HOLMES, C. C. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Anal.* **1** 145–168.

HUANG, A. and WAND, M. P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Anal.* **8** 439–452.

ILSE, M., TOMCZAK, J. and WELLING, M. (2018). Attention-based deep multiple instance learning. In *Proceedings of the* 35*th International Conference on Machine Learning* 2127–2136.

IWAHORI, K., KAKARLA, S., VELASQUEZ, M. P., YU, F., YI, Z., GERKEN, C., SONG, X.-T. and GOTTSCHALK, S. (2015). Engager T cells: A new class of antigen-specific T cells that redirect bystander T cells. *Mol. Ther.* **23** 171–178.

LAMBRECHTS, D., WAUTERS, E., BOECKX, B., AIBAR, S., NITTNER, D., BURTON, O., BASSEZ, A., DECALUWÉ, H., PIRCHER, A. et al. (2018). Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24** 1277–1289.

LIU, J., LICHTENBERG, T., HOADLEY, K. A., POISSON, L. M., LAZAR, A. J., CHERNIACK, A. D., KOVATICH, A. J., BENZ, C. C., LEVINE, D. A. et al. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173** 400–416.

LU, T., ZHANG, Z., ZHU, J., WANG, Y., JIANG, P., XIAO, X., BERNATCHEZ, C., HEYMACH, J. V., GIBBONS, D. L. et al. (2021). Deep learning-based prediction of the T cell receptor–antigen binding specificity. *Nat. Mach. Intell.* 1–12.

MARON, O. and LOZANO-PÉREZ, T. (1998). A framework for multiple-instance learning. In *Advances in Neural Information Processing Systems* 570–576.

MIAO, D., MARGOLIS, C. A., GAO, W., VOSS, M. H., LI, W., MARTINI, D. J., NORTON, C., BOSSÉ, D., WANKOWICZ, S. M. et al. (2018). Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science* **359** 801–806.

NEWTON, M. A., NOUEIRY, A., SARKAR, D. and AHLQUIST, P. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* **5** 155–176. https://doi.org/10.1093/biostatistics/5.2.155

PARK, S., WANG, X., LIM, J., XIAO, G., LU, T. and WANG, T. (2020). Bayesian multiple instance regression for modeling immunogenic neoantigens. *Stat. Methods Med. Res.* **29** 3032–3047.

PENDERGAST, J. F., GANGE, S. J., NEWTON, M. A., LINDSTROM, M. J., PALTA, M. and FISHER, M. R. (1996). A survey of methods for analyzing clustered binary response data. *International Statistical Review/Revue Internationale de Statistique* 89–118.

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.

RAY, S. and CRAVEN, M. (2005). Supervised versus multiple instance learning: An empirical comparison. In *Proceedings of the* 22*nd International Conference on Machine Learning* 697–704.

RAY, S. and PAGE, D. (2001). Multiple instance regression. In *Proceedings of the* 18*th Internatinal Conference on Machine Learning* 425–432.

SATO, Y., YOSHIZATO, T., SHIRAISHI, Y., MAEKAWA, S., OKUNO, Y., KAMURA, T., SHIMAMURA, T., SATO-OTSUBO, A., NAGAE, G. et al. (2013). Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.* **45** 860–867.

SINGHI, A. D., KOAY, E. J., CHARI, S. T. and MAITRA, A. (2019). Early detection of pancreatic cancer: Opportunities and challenges. *Gastroenterology* **156** 2024–2040. https://doi.org/10.1053/j.gastro.2019.01.259

THE CANCER GENOME ATLAS RESEARCH NETWORK (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499** 43–49.

VANDERLUGT, C. L. and MILLER, S. D. (2002). Epitope spreading in immune-mediated diseases: Implications for immunotherapy. *Nat. Rev., Immunol.* **2** 85–95. https://doi.org/10.1038/nri724

WANG, J. and ZUCKER, J.-D. (2000). Solving multiple-instance problem: A lazy learning approach. In *Proceedings of the* 17*th International Conference on Machine Learning* 1119–1126.

WANG, T., LU, R., KAPUR, P., JAISWAL, B. S., HANNAN, R., ZHANG, Z., PEDROSA, I., LUKE, J. J., ZHANG, H., GOLDSTEIN, L. D. et al. (2018). An empirical approach leveraging tumorgrafts to dissect the tumor microenvironment in renal cell carcinoma identifies missing link to prognostic inflammatory factors. *Cancer Discov.* **8** 1142–1155.

WANG, X., YAN, Y., TANG, P., BAI, X. and LIU, W. (2018). Revisiting multiple instance neural networks. *Pattern Recognit.* **74** 15–24.

WANG, Z., RADOSAVLJEVIC, V., HAN, B., OBRADOVIC, Z. and VUCETIC, S. (2008). Aerosol optical depth prediction from satellite observations by multiple instance regression. In *Proceedings of the* 2008 *SIAM International Conference on Data Mining* 165–176.

WHITESIDE, S. K., SNOOK, J. P., WILLIAMS, M. A. and WEIS, J. J. (2018). Bystander T cells: A balancing act of friends and foes. *Trends Immunol.* **39** 1021–1035. https://doi.org/10.1016/j.it.2018.10.003

XIONG, D., PARK, S., LIM, J., WANG, T. and WANG, X. (2024). Supplement to “Bayesian multiple instance classification based on hierarchical probit regression.” https://doi.org/10.1214/23-AOAS1780SUPPA, https://doi.org/10.1214/23-AOAS1780SUPPB

XIONG, D., ZHANG, Z., WANG, T. and WANG, X. (2021). A comparative study of multiple instance learning methods for cancer detection using T-cell receptor sequences. *Comput. Struct. Biotechnol. J.* **19** 3255.

ZHANG, J., MARSZAŁEK, M., LAZEBNIK, S. and SCHMID, C. (2007). Local features and kernels for classification of texture and object categories: A comprehensive study. *Int. J. Comput. Vis.* **73** 213–238.

ZHANG, Q. and GOLDMAN, S. A. (2002). EM-DD: An improved multiple-instance learning technique. In *Advances in Neural Information Processing Systems* 1073–1080.

ZHANG, Z., XIONG, D., WANG, X., LIU, H. and WANG, T. (2021). Mapping the functional landscape of T cell receptor repertoires by single-T cell transcriptomics. *Nat. Methods* **18** 92–99.

ZHOU, Z.-H., SUN, Y.-Y. and LI, Y.-F. (2009). Multi-instance learning by treating instances as non-iid samples. In *Proceedings of the* 26*th International Conference on Machine Learning* 1249–1256.

ZHOU, Z.-H. and ZHANG, M.-L. (2007). Solving multi-instance problems with classifier ensemble based on constructive clustering. *Knowl. Inf. Syst.* **11** 155–170.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 100–124
https://doi.org/10.1214/23-AOAS1781

ESTIMATING FIBER ORIENTATION DISTRIBUTION WITH APPLICATION TO STUDY BRAIN LATERALIZATION USING HCP D-MRI DATA

BY SEUNGYONG HWANG[a], THOMAS C. M. LEE[b], DEBASHIS PAUL[c] AND JIE PENG[d]

Department of Statistics, University of California, Davis, [a]*syhwang@ucdavis.edu,* [b]*tcmlee@ucdavis.edu,* [c]*debpaul@ucdavis.edu,* [d]*jiepeng@ucdavis.edu*

Diffusion-weighted magnetic resonance imaging (D-MRI) is an in vivo and noninvasive imaging technology for characterizing tissue microstructure in biological samples. A major application of D-MRI is for white matter fiber tract reconstruction in brains. It begins by estimating the water molecule movements (serving as proxies for fiber directions) in the brain voxels and then combines the results to form fiber tracts. The voxel-level fiber direction information can be modeled by a *fiber orientation distribution* (FOD) function, and in this paper, we propose a computationally scalable FOD estimator, the *blockwise James–Stein* (*BJS*) estimator. We then apply *BJS* to the D-MRI data from the *Human Connectome Project (HCP)* to study brain lateralization, an important topic in neuroscience. Specifically, we focus on the association between lateralization of the *superior longitudinal fasciculus* (SLF)—a major association tract and handedness. For each subject from the HCP data, we extract voxel-level directional information by *BJS* and then reconstruct the SLF in each brain hemisphere through a tractography algorithm. Finally, we derive a lateralization score that quantifies hemispheric asymmetry of the reconstructed SLF. We then relate this lateralization score to gender and handedness through an ANOVA model, where significant handedness effects are found. The results indicate that the SLF lateralization is likely to be different in right-handed and left-handed individuals. Codes and example scripts for both synthetic experiments and HCP data application can be found at https://github.com/vic-dragon/BJS.

REFERENCES

ALEXANDER, A. L. (2010). Deterministic white matter tractography. In *Diffusion MRI*: *Theory*, *Methods*, *and Applications* (D. K. Jones, ed.) 201–213. Oxford Univ. Press, London.

AMEMIYA, K., NAITO, E. and TAKEMURA, H. (2021). Age dependency and lateralization in the three branches of the human superior longitudinal fasciculus. *Cortex* **139** 116–133.

BASSER, P. J., PAJEVIC, S., PIERPAOLI, C., DUDA, J. and ALDROUBI, A. (2000). In vivo fiber tractography using DT-MRI data. *Magn*. *Reson*. *Med*. **44** 625–632. https://doi.org/10.1002/1522-2594(200010)44:4<625::aid-mrm17>3.0.co;2-o

BUDISAVLJEVIC, S., CASTIELLO, U. and BEGLIOMINI, C. (2021). Handedness and white matter networks. *Neuroscientist* **27** 88–103. https://doi.org/10.1177/1073858420937657

BUDISAVLJEVIC, S., DELL'ACQUA, F., ZANATTO, D., BEGLIOMINI, C., MIOTTO, D., MOTTA, R. and CASTIELLO, U. (2017). Asymmetry and structure of the fronto-parietal networks underlie visuomotor processing in humans. *Cereb*. *Cortex* **27** 1532–1544. https://doi.org/10.1093/cercor/bhv348

CAI, T. T. (1999). Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann*. *Statist*. **27** 898–924. MR1724035 https://doi.org/10.1214/aos/1018031262

CAI, T. T., LOW, M. G. and ZHAO, L. H. (2009). Sharp adaptive estimation by a blockwise method. *J*. *Nonparametr*. *Stat*. **21** 839–850. MR2572586 https://doi.org/10.1080/10485250903142663

CARMICHAEL, O., CHEN, J., PAUL, D. and PENG, J. (2013). Diffusion tensor smoothing through weighted Karcher means. *Electron*. *J*. *Stat*. **7** 1913–1956. MR3084676 https://doi.org/10.1214/13-EJS825

CATANI, M. (2010). The functional anatomy of white matter: From postmortem dissections to in vivo virtual tractography. In *Diffusion MRI*: *Theory*, *Methods*, *and Applications* (D. K. Jones, ed.) 5–18. Oxford Univ. Press, London.

Key words and phrases. Neuroimaging, tractography, spherical convolution, James–Stein estimator.

CATANI, M., ALLIN, M. P. G., HUSAIN, M., PUGLIESE, L., MESULAM, M. M., MURRAY, R. M. and JONES, D. K. (2007). Symmetries in human brain language pathways correlate with verbal recall. *Proc. Natl. Acad. Sci. USA* **104** 17163–17168.

CAVALIER, L. and TSYBAKOV, A. (2002). Sharp adaptation for inverse problems with random noise. *Probab. Theory Related Fields* **123** 323–354. MR1918537 https://doi.org/10.1007/s004400100169

CAVALIER, L. and TSYBAKOV, A. B. (2001). Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation *Math. Methods Statist.* **10** 247–282. MR1867161

DE GROOT, M., VERNOOIJ, M. W., KLEIN, S., IKRAM, M. A., VOS, F. M., SMITH, S. M., NIESSEN, W. J. and ANDERSSON, J. L. R. (2013). Improving alignment in tract-based spatial statistics: Evaluation and optimization of image registration. *NeuroImage* **76** 400–411.

DESCOTEAUX, M., ANGELINO, E., FITZGIBBONS, S. and DERICHE, R. (2006). Apparent diffusion coefficients from high angular resolution diffusion imaging: Estimation and applications. *Magn. Reson. Med.* **56** 395–410.

DESCOTEAUX, M., ANGELINO, E., FITZGIBBONS, S. and DERICHE, R. (2007). Regularized, fast and robust analytical Q-ball imaging. *Magn. Reson. Med.* **58** 497–510.

ESSEN, D. C. V., SMITH, S. M., BARCH, D. M., BEHRENS, T. E. J., YACOUB, E. and UGURBIL, K. (2013). The Wu-Minn human connectome project: An overview. *NeuroImage* **80** 62–79.

GHARABAGHI, A., KUNATH, F., ERB, M., SAUR, R., HECKL, S., TATAGIBA, M., GRODD, W. and KARNATH, H.-O. (2009). Perisylvian white matter connectivity in the human right hemisphere. *BMC Neurosci.* **10** 15. https://doi.org/10.1186/1471-2202-10-15

GLASSER, M. F., SOTIROPOULOS, S. N., WILSON, J. A., COALSON, T. S., FISCHL, B., ANDERSSON, J. L., XU, J., JBABDI, S., WEBSTER, M. et al. (2013). The minimal preprocessing pipelines for the human connectome project. *NeuroImage* **80** 105–124.

GOLDENSHLUGER, A. and TSYBAKOV, A. (2001). Adaptive prediction and estimation in linear regression with infinitely many parameters. *Ann. Statist.* **29** 1601–1619. MR1891740 https://doi.org/10.1214/aos/1015345956

GUDBJARTSSON, H. and PATZ, S. (1995). The Rician distribution of noisy MRI data. *Magn. Reson. Med.* **34** 910–914. https://doi.org/10.1002/mrm.1910340618

HEALY, D. M. JR., HENDRIKS, H. and KIM, P. T. (1998). Spherical deconvolution. *J. Multivariate Anal.* **67** 1–22. MR1659108 https://doi.org/10.1006/jmva.1998.1757

HOSEY, T., WILLIAMS, G. and ANSORGE, R. (2005). Inference of multiple fiber orientations in high angular resolution diffusion imaging. *Magn. Reson. Med.* **54** 1480–1489.

HOUSTON, J., ALLENDORFER, J., NENERT, R., GOODMAN, A. M. and SZAFLARSKI, J. P. (2019). White matter language pathways and language performance in healthy adults across ages. *Front. Neurosci.* **13** 1185. https://doi.org/10.3389/fnins.2019.01185

HOWELLS, H., THIEBAUT DE SCHOTTEN, M., DELL'ACQUA, F., BEYH, A., ZAPPALÀ, G., LESLIE, A., SIMMONS, A., MURPHY, D. G. and CATANI, M. (2018). Frontoparietal tracts linked to lateralized hand preference and manual specialization. *Cereb. Cortex* **28** 2482–2494.

HUA, K., ZHANG, J., WAKANA, S., JIANG, H., LI, X., REICH, D. S., CALABRESI, P. A., PEKAR, J. J., VAN ZIJL, P. C. M. et al. (2008). Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. *NeuroImage* **39** 336–347.

HWANG, S., LEE, T. C. M., PAUL, D. and PENG, J. (2024). Supplement to "Estimating Fiber Orientation Distribution with Application to Study Brain Lateralization using HCP D-MRI Data." https://doi.org/10.1214/23-AOAS1781SUPPA, https://doi.org/10.1214/23-AOAS1781SUPPB

JENKINSON, M., BANNISTER, P., BRADY, M. and SMITH, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* **17** 825–841. https://doi.org/10.1016/s1053-8119(02)91132-8

JENKINSON, M., BECKMANN, C. F., BEHRENS, T. E. J., WOOLRICH, M. W. and SMITH, S. M. (2012). FSL. *NeuroImage* **62** 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

JONES, D. K. (2010). *Diffusion MRI: Theory, Methods, and Applications*. Oxford Univ. Press, London.

JONES, D. K., KNÖSCHE, T. R. and TURNER, R. (2013). White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion MRI. *NeuroImage* **73** 239–254. https://doi.org/10.1016/j.neuroimage.2012.06.081

LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. MR1805785 https://doi.org/10.1214/aos/1015957395

LE BIHAN, D., MANGIN, J. F., POUPON, C., CLARK, A. C., PAPPATA, S., MOLKO, N. and CHABRIAT, H. (2001). Diffusion tensor imaging: Concepts and applications. *J. Magn. Reson. Imaging* **13** 534–546.

LÓPEZ-VICENTE, M., LAMBALLAIS, S., LOUWEN, S., HILLEGERS, M., TIEMEIER, H., MUETZEL, R. L. and WHITE, T. (2021). White matter microstructure correlates of age, sex, handedness and motor ability in a population-based sample of 3031 school-age children. *NeuroImage* **227** 117643.

MAKRIS, N., KENNEDY, D. N., MCINERNEY, S., SORENSEN, A. G., WANG, R., CAVINESS JR, V. S. and PANDYA, D. N. (2004). Segmentation of subcomponents within the superior longitudinal fascicle in humans: A quantitative, in vivo, DT-MRI study. *Cereb. Cortex* **15** 854–869.

MCCARTHY, P. (2020). FSLeyes 0.34.0. Available at https://doi.org/10.5281/zenodo.3937147.

MORI, S. (2007). *Introduction to Diffusion Tensor Imaging*. Elsevier, Amsterdam.

MUSCHELLI, J. (2018). neurohcp: Human 'Connectome' Project Interface. R package version 0.8.1.

MUSCHELLI, J., SWEENEY, E., LINDQUIST, M. and CRAINICEANU, C. (2015). fslr: Connecting the FSL software with R. *R J.* **7** 163–175.

NIMSKY, C., GANSLANDT, O. and FAHLBUSCH, R. (2006). Implementation of fiber tract navigation. *Neurosurgery* **58** 292–303.

RAO, S., IBRAHIM, J. G., CHENG, J., YAP, P.-T. and ZHU, H. (2016). SR-HARDI: Spatially regularizing high angular resolution diffusion imaging. *J. Comput. Graph. Statist.* **25** 1195–1211. MR3572036 https://doi.org/10.1080/10618600.2015.1105750

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

SEIZEUR, R., MAGRO, E., PRIMA, S., WIEST-DAESSLÉ, N., MAUMET, C. and MORANDI, X. (2014). Corticospinal tract asymmetry and handedness in right- and left-handers by diffusion tensor tractography. *Surg. Radiol. Anat.* **36** 111–124. https://doi.org/10.1007/s00276-013-1156-7

SMITH, S. M., JENKINSON, M., JOHANSEN-BERG, H., RUECKERT, D., NICHOLS, T. E., MACKAY, C. E., WATKINS, K. E., CICCARELLI, O., CADER, M. Z. et al. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage* **31** 1487–1505. https://doi.org/10.1016/j.neuroimage.2006.02.024

SPORNS, O. (2011). *Networks of the Brain*. MIT Press, Cambridge.

SZAFLARSKI, J. P., RAJAGOPAL, A., ALTAYE, M., BYARS, A. W., JACOLA, L. and SCHMITHORST, V. (2012). Left-handedness and language lateralization in children. *Brain Res.* **1433** 85–97.

TOURNIER, J., CALAMANTE, F. and CONNELLY, A. (2007). Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage* **35** 1459–1472. https://doi.org/10.1016/j.neuroimage.2007.02.016

TOURNIER, J., CALAMANTE, F., GADIAN, D. G. and CONNELLY, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* **23** 1176–1185. https://doi.org/10.1016/j.neuroimage.2004.07.037

TUCH, D. S. (2002). Diffusion MRI of complex tissue structure Ph.D. thesis Massachusetts Institute of Technology.

TUCH, D. S. (2004). Q-ball imaging. *Magn. Reson. Med.* **52** 1358–1372.

TUCH, D. S., REESE, T. G., WIEGELL, M. R., MAKRIS, N., BELLIVEAU, J. W. and WEDEEN, V. J. (2002). High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magn. Reson. Med.* **48** 577–582. https://doi.org/10.1002/mrm.10268

WAKANA, S., CAPRIHAN, A., PANZENBOECK, M. M., FALLON, J. H., PERRY, M., GOLLUB, R. L., HUA, K., ZHANG, J., JIANG, H. et al. (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter. *NeuroImage* **36** 630–644.

WONG, R. K. W., LEE, T. C. M., PAUL, D., PENG, J. and INITIATIVE, A. D. N. (2016). Fiber direction estimation, smoothing and tracking in diffusion MRI. *Ann. Appl. Stat.* **10** 1137–1156. MR3553216 https://doi.org/10.1214/15-AOAS880

WOOLRICH, M. W., JBABDI, S., PATENAUDE, B., CHAPPELL, M., MAKNI, S., BEHRENS, T., BECKMANN, C., JENKINSON, M. and SMITH, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage* **45** S173–S186.

YAN, H., CARMICHAEL, O., PAUL, D., PENG, J. and ADNI (2018). Estimating fiber orientation distribution from diffusion MRI with spherical needlets. *Med. Image Anal.* **46** 57–72.

YEH, F.-C., VERSTYNEN, T. D., WANG, Y., FERNÁNDEZ-MIRANDA, J. C. and TSENG, W.-Y. I. (2013). Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PLoS ONE* **8**.

YEH, F. C., WEEDEN, V. J. and TSENG, W. Y. I. (2010). Generalized q-sampling imaging. *IEEE Trans. Med. Imag.* **29** 1626–1635.

ZHANG, Y., BRADY, M. and SMITH, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imag.* **20** 45–57. https://doi.org/10.1109/42.906424

ZHANG, Z., DESCOTEAUX, M., ZHANG, J., GIRARD, G., CHAMBERLAND, M., DUNSON, D., SRIVASTAVA, A. and ZHU, H. (2018). Mapping population-based structural connectomes. *NeuroImage* **172** 130–145.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 125–143
https://doi.org/10.1214/23-AOAS1782

USING SIMULTANEOUS REGRESSION CALIBRATION TO STUDY THE EFFECT OF MULTIPLE ERROR-PRONE EXPOSURES ON DISEASE RISK UTILIZING BIOMARKERS DEVELOPED FROM A CONTROLLED FEEDING STUDY

BY YIWEN ZHANG[1,a], RAN DAI[2,b], YING HUANG[3,d], ROSS PRENTICE[3,e] AND CHENG ZHENG[2,c]

[1]*Zilber School of Public Health, University of Wisconsin-Milwaukee,* [a]*yzhang693@outlook.com*
[2]*Department of Biostatistics, University of Nebraska Medical Center,* [b]*ran.dai@unmc.edu,* [c]*cheng.zheng@unmc.edu*
[3]*Public Health Science Division, Fred Hutchinson Cancer Center,* [d]*yhuang@fhcrc.org,* [e]*rprentic@whi.org*

Systematic measurement error in self-reported data creates important challenges in association studies between dietary intakes and chronic disease risks, especially when multiple dietary components are studied jointly. The joint regression calibration method has been developed for measurement error correction when objectively measured biomarkers are available for all dietary components of interest. Unfortunately, objectively measured biomarkers are only available for very few dietary components, which limits the application of the joint regression calibration method. Recently, for single dietary components, controlled feeding studies have been performed to develop new biomarkers for many more dietary components. However, it is unclear whether the biomarkers separately developed for single dietary components are valid for joint calibration. In this paper we show that biomarkers developed for single dietary components cannot be used for joint regression calibration. We propose new methods to utilize controlled feeding studies to develop valid biomarkers for joint regression calibration to estimate the association between multiple dietary components simultaneously with the disease of interest. Asymptotic distribution theory for the proposed estimators is derived. Extensive simulations are performed to study the finite sample performance of the proposed estimators. We apply our methods to examine the joint effects of sodium and potassium intakes on cardiovascular disease incidence using the Women's Health Initiative cohort data. We identify positive associations between sodium intake and cardiovascular diseases as well as negative associations between potassium intake and cardiovascular disease.

REFERENCES

ADAMS, K. F., SCHATZKIN, A., HARRIS, T. B., KIPNIS, V., MORRIS, T. and BALLARD-BARBASH, R. (2006). Overweight, obesity and mortality in a large prospective cohort of persons 50 to 71 years old. *N. Engl. J. Med.* **355** 763–778.

BARTLETT, J. W. and KEOGH, R. H. (2018). Bayesian correction for covariate measurement error: A frequentist evaluation and comparison with regression calibration. *Stat. Methods Med. Res.* **27** 1695–1708.

CARROLL, R. J., RUPPERT, D. and STEFANSKI, L. A. (1995). Measurement error in nonlinear models Chapman and Hall London.

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models*: *A Modern Perspective*. CRC Press, US.

COOK, N. R., OBARZANEK, E., CUTLER, J. A., BURING, J. E., REXRODE, K. M., KUMANYIKA, S. K., APPEL, L. J. and WHELTON, P. K. (2009). Joint effects of sodium and potassium intake on subsequent cardiovascular disease: The trials of hypertension prevention follow-up study. *Arch. Intern. Med.* **169** 32–40.

. WORLD CANCER RESEARCH FUND/AMERICAN INSTITUTE FOR CANCER RESEARCH (WCRF/AICR) (2007). *Food*, *Nutrition and the Prevention of Cancer*: *A Global Perspective*. American Institute for Cancer Research, Washington, DC.

Key words and phrases. Measurement error, regression calibration, feeding study, biomarker, cardiovascular disease.

FREEDMAN, L. S., SCHATZKIN, A., MIDTHUNE, D. and KIPNIS, V. (2011). Dealing with dietary measurement error in nutritional cohort studies. *J. Natl. Cancer Inst.* **103** 1086–1092. https://doi.org/10.1093/jnci/djr189

GELEIJNSE, J. M., KOK, F. J. and GROBBEE, D. E. (2004). Impact of dietary and lifestyle factors on the prevalence of hypertension in Western populations. *Eur. J. Public Health* **14** 235–239. https://doi.org/10.1093/eurpub/14.3.235

GLATZ, N., CHAPPUIS, A., CONEN, D., ERNE, P., PÉCHÈRE-BERTSCHI, A., GUESSOUS, I., FOGNA, V., GABUTTI, L., MUGGLI, F. et al. (2017). Associations of sodium, potassium and protein intake with blood pressure and hypertension in Switzerland. *Swiss Med. Wkly.* **147** w14411.

HU, C. and LIN, D. Y. (2002). Cox regression with covariate measurement error. *Scand. J. Stat.* **29** 637–655.

HUANG, Y., VAN HORN, L., TINKER, L. F., NEUHOUSER, M. L., CARBONE, L., MOSSAVAR-RAHMANI, Y., THOMAS, F. and PRENTICE, R. L. (2014). Measurement error corrected sodium and potassium intake estimation using 24-hour urinary excretion. *Hypertens.* **63** 238–244.

HUANG, Y. and WANG, C. Y. (2000). Cox regression with accurate covariate unascertainable: A nonparametric-correction approach. *J. Amer. Statist. Assoc.* **45** 1209–1219.

HUANG, Y., ZHENG, C., TINKER, L., NEUHOUSER, M. and PRENTICE, R. (2022). Biomarker-based methods and study designs to calibrate dietary intake for assessing diet-disease associations. *J. Nutr.* **152** 899–906.

KORZUN, W. and MILLER, W. (1987). Sodium and potassium. In *Methods in Clinical Chemistry* (A. Pesce and K. L, eds.) p. 86. CV Mosby, St. Louis MO.

LAMPE, J. W., HUANG, Y., NEUHOUSER, M. L., TINKER, L. F., SONG, X., SCHOELLER, D. A., KIM, S., RAFTERY, D., DI, C. et al. (2017). Dietary biomarker evaluation in a controlled feeding study in women from the women's health initiative cohort. *Am. J. Clin. Nutr.* **105** 466–475.

LI, Y. and RYAN, L. (2006b). Inference on survival data with covariate measurement error-an imputationapproach. *Scand. J. Stat.* **33** 169–190.

MIZÉHOUN-ADISSODA, C., HOUINATO, D., HOUEHANOU, C., CHIANEA, T., DALMAY, F., BIGOT, A., ABOYANS, V., PREUX, P.-M., BOVET, P. et al. (2017). Dietary sodium and potassium intakes: Data from urban and rural areas. *Nutr.* **33** 35–41.

O'DONNELL, M., MENTE, A., RANGARAJAN, S., MCQUEEN, M. J., WANG, X., LIU, L., YAN, H., LEE, S. F., MONY, P. et al. (2014). Urinary sodium and potassium excretion, mortality, and cardiovascular events. *N. Engl. J. Med.* **371** 612–623.

PAERATAKUL, S., POPKIN, B. M., KOHLMEIER, L., HERTZ-PICCIOTTO, I., GUO, X. and EDWARDS, L. J. (1998). Measurement error in dietary data: Implications for the epidemiologic study of the diet-disease relationship. *Eur. J. Clin. Nutr.* **52** 722–727.

PRENTICE, R., PETTINGER, M., NEUHOUSER, M., RAFTERY, D., ZHENG, C., GOWDA, N., HUANG, Y., TINKER, L., HOWARD, B. et al. (2021). Biomarker-calibrated macronutrient intake and chronic disease risk among postmenopausal women. *J. Nutr.* **151** 2330–2341.

PRENTICE, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69** 331–342.

PRENTICE, R. L., ARAGAKI, A. K., VAN HORN, L., THOMSON, C. A., TINKER, L. F., MANSON, J. E., MOSSAVAR-RAHMANI, Y., HUANG, Y., ZHENG, C. et al. (2022). Mortality associated with healthy eating index components and an empirical-scores healthy eating index in a cohort of postmenopausal women. *J. Nutr.* **152** 2493–2504.

PRENTICE, R. L., HUANG, Y., NEUHOUSER, M. L., MANSON, J. E., MOSSAVAR-RAHMANI, Y., THOMAS, F., TINKER, L. F., ALLISON, M., JOHNSON, K. C. et al. (2017). Associations of biomarker-calibrated sodium and potassium intakes with cardiovascular disease risk among postmenopausal women. *Amer. J. Epidemiol.* **186** 1035–1043.

PRENTICE, R. L., MOSSAVAR-RAHMANI, Y., HUANG, Y., HORN, L. V., BERESFORD, S. A. A., CAAN, B., TINKER, L., SCHOELLER, D., BINGHAM, S. et al. (2011). Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *Amer. J. Epidemiol.* **174** 591–603.

PRENTICE, R. L., TINKER, L. F., HUANG, Y. and NEUHOUSER, M. L. (2013). Calibration of self-reported dietary measures using biomarkers: An approach to enhancing nutritional epidemiology reliability. *Curr. Atheroscl. Rep.* **15** 353.

ROSNER, B., SPIEGELMAN, D. and WILLETT, W. C. (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: The case of multiple covariates measured with error. *Amer. J. Epidemiol.* **132** 734–745.

SHAW, P. A. and PRENTICE, R. L. (2012). Hazard ratio estimation for biomarker-calibrated dietary exposures. *Biometrics* **68** 397–407.

SONG, X. and HUANG, X. (2005). On corrected score approach for proportional hazards model with covariate-measurement error. *Biometrics* **61** 702–714.

WANG, C. Y., HSU, L., FENG, Z. D. and PRENTICE, R. L. (1997). Regression calibration in failure time regression. *Biometrics* **53** 131–145.

XI, L., HAO, Y.-C., LIU, J., WANG, W., WANG, M., LI, G.-Q., QI, Y., ZHAO, F., XIE, W.-X. et al. (2015). Associations between serum potassium and sodium levels and risk of hypertension: A community-based cohort study. *J. Geriatr. Cardiol.* **12** 119–126.

YAN, Y. and YI, G. Y. (2015b). A corrected profile likelihood method for survival data with covariate measurement error under the Cox model. *Canad. J. Statist.* **43** 454–480.

ZHANG, Y., DAI, R., HUANG, Y., PRENTICE, R. and ZHENG, C. (2024). Supplement to "Using simultaneous regression calibration to study the effect of multiple error-prone exposures on disease risk utilizing biomarkers developed from a controlled feeding study." https://doi.org/10.1214/23-AOAS1782SUPP

ZHENG, C., BERESFORD, S. A. A., HORN, L. V., TINKER, L. F., THOMSON, C. A., NEUHOUSER, M. L., DI, C., MANSON, J. E., MOSSAVAR-RAHMANI, Y. et al. (2014). Simultaneous association of total energy consumption and activity-related energy expenditure with cardiovascular disease, cancer, and diabetes risk among postmenopausal women. *Amer. J. Epidemiol.* **180** 526–535.

ZHENG, C., ZHANG, Y., HUANG, Y. and PRENTICE, R. (2023). Using controlled feeding study for biomarker development in regression calibration for disease association estimation. *Stat. Biosci.* **15** 57–113.

ZUCKER, D. M. (2005b). A pseudo-partial likelihood method for semiparametric survival regression withcovariate errors. *J. Amer. Statist. Assoc.* **100** 1264–1277.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 144–158
https://doi.org/10.1214/23-AOAS1783

A QUANTITATIVE LINGUISTIC ANALYSIS OF A CANCER ONLINE HEALTH COMMUNITY WITH A SMOOTH LATENT SPACE MODEL

BY MENGQUE LIU[1,a], XINYAN FAN[2,b] AND SHUANGGE MA[3,c]

[1]*School of Journalism and New Media, Xi'an Jiaotong University,* [a]*mengqueliu@xjtu.edu.cn*
[2]*Center for Applied Statistics and School of Statistics, Renmin University of China,* [b]*1031820039@qq.com*
[3]*Department of Biostatistics, Yale University,* [c]*shuangge.ma@yale.edu*

Online health communities (OHCs) provide free, open, and well-resourced platforms for patients, family members, and others to discuss illnesses, express feelings, and connect with others. Linguistic analysis of OHC posts can assist in better understanding disease conditions as well as monitoring the emotional and mental status of patients and those who are closely related. Many existing OHC linguistic analyses are limited by focusing on individual words. There are a handful of cooccurrence network analyses, which have multiple methodological limitations. In this article we analyze posts that are publicly available at the LUNGevity Foundation's Lung Cancer Support Community (LCSC). The analyzed data contains 21,028 posts published between April 2018 and February 2022. For word cooccurrence network analysis, we develop a two-part latent space model, which advances from the existing ones by accommodating network weights. Further, we consider the scenario where there are change points in time, networks remain the same between two change points but differ on the two sides of a change point, and the number and locations of change points are unknown. A penalized fusion approach is developed to data-dependently determine change points and estimate networks. In data analysis multiple change points are identified, which reflect significant changes in lung cancer patients' and their close affiliates' emotional/mental status and mostly align with the changes in COVID-19. The obtained network structures and other findings are also sensible.

REFERENCES

AUSTENFELD, J. L. and STANTON, A. L. (2004). Coping through emotional approach: A new look at emotion, coping, and health-related outcomes. *J. Pers.* **72** 1335–1364.

BÄUERLE, A., MUSCHE, V., SCHMIDT, K., SCHWEDA, A., FINK, M., WEISMÜLLER, B., KOHLER, H., HERRMANN, K., TEWES, M. et al. (2021). Mental health burden of German cancer patients before and after the outbreak of Covid-19: Predictors of mental health impairment. *Int. J. Environ. Res. Public Health* **18** 2318.

BLEAKLEY, K. and VERT, J.-P. (2011). The group fused lasso for multiple change-point detection. ArXiv preprint. Available at arXiv:1106.4199.

BOLLEN, K. A. and CURRAN, P. J. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociol. Methods Res.* **32** 336–383. MR2037798 https://doi.org/10.1177/0049124103260222

BOLLEN, K. A. and CURRAN, P. J. (2006). *Latent Curve Models*: *A Structural Equation Perspective*. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2184502

BOYD, S., PARIKH, N., CHU, E., PELEATO, B. and ECKSTEIN, J. (2010). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3** 1–122.

CHAVES, A. L. F., CASTRO, A. F., MARTA, G. N., JUNIOR, G. C., FERRIS, R. L., GIGLIO, R. E., GOLUSINSKI, W., GORPHE, P., HOSAL, S. et al. (2020). Emergency changes in international guidelines on treatment for head and neck cancer patients during the Covid-19 pandemic. *Oral Oncol.* **107** 104734.

DEHDARIRAD, T. and FREER, J. (2021). Is there alignment amongst scientific literature, news media and patient forums regarding topics?: A study of breast and lung cancer. *Online Inf. Rev.* **45** 983–999.

GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2010). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.

Key words and phrases. Cancer, cooccurrence network, online health community, quantitative linguistic analysis, smooth latent space model.

GOLDER, S. A. and MACY, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* **333** 1878–1881. https://doi.org/10.1126/science.1202775

HARTLEY, C. A. and PHELPS, E. A. (2012). Anxiety and decision-making. *Biol. Psychiatry* **72** 113–118. https://doi.org/10.1016/j.biopsych.2011.12.027

HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262 https://doi.org/10.1198/016214502388618906

HOFMANN, S. G., MOORE, P. M., GUTNER, C. and WEEKS, J. W. (2012). Linguistic correlates of social anxiety disorder. *Cogn. Emot.* **26** 720–726. https://doi.org/10.1080/02699931.2011.602048

KE, Y., LI, J. and ZHANG, W. (2016). Structure identification in panel data analysis. *Ann. Statist.* **44** 1193–1233. MR3485958 https://doi.org/10.1214/15-AOS1403

KE, Z. T., FAN, J. and WU, Y. (2015). Homogeneity pursuit. *J. Amer. Statist. Assoc.* **110** 175–194. MR3338495 https://doi.org/10.1080/01621459.2014.892882

KUDERER, N. M., CHOUEIRI, T. K., SHAH, D. P., SHYR, Y., RUBINSTEIN, S. M., RIVERA, D. R., SHETE, S., HSU, C.-Y., DESAI, A. et al. (2020). Clinical impact of Covid-19 on patients with cancer (CCC19): A cohort study. *Lancet* **395** 1907–1918.

LIN, K., SHARPNACK, J., RINALDO, A. and TIBSHIRANI, R. J. (2017). A sharp error analysis for the fused lasso, with application to approximate changepoint screening. In *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*. *NIPS'*17 6887–6896. Curran Associates, Red Hook, NY.

LIU, M., FAN, X. and MA, S. (2024). Supplement to "A quantitative linguistic analysis of a cancer online health community with a smooth latent space model." https://doi.org/10.1214/23-AOAS1783SUPP

MA, Z., MA, Z. and YUAN, H. (2020). Universal latent space model fitting for large networks with edge covariates. *J. Mach. Learn. Res.* **21** Paper No. 4. MR4071187 https://doi.org/10.1109/tnnls.2020.3010690

MEHLER, A., LÜCKING, A., BANISCH, S., BLANCHARD, P. and JOB, B. (2016). *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*. Springer, Berlin.

NAUSHEEN, B., GIDRON, Y., PEVELER, R. and MOSS-MORRIS, R. (2009). Social support and cancer progression: A systematic review. *J. Psychosom. Res.* **67** 403–415. https://doi.org/10.1016/j.jpsychores.2008.12.012

PALOWITCH, J., BHAMIDI, S. and NOBEL, A. B. (2017). Significance-based community detection in weighted networks. *J. Mach. Learn. Res.* **18** Paper No. 188. MR3827076

PEH, C. X., KUA, E. H. and MAHENDRAN, R. (2016). Hope, emotion regulation, and psychosocial well-being in patients newly diagnosed with cancer. *Support Care Cancer* **24** 1955–1962. https://doi.org/10.1007/s00520-015-2989-x

RESHEF, D. N., RESHEF, Y. A., FINUCANE, H. K., GROSSMAN, S. R., MCVEAN, G., TURNBAUGH, P. J., LANDER, E. S., MITZENMACHER, M. and SABETI, P. C. (2011). Detecting novel associations in large data sets. *Science* **334** 1518–1524.

ROGERS, J. P., CHESNEY, E., OLIVER, D., POLLAK, T. A., MCGUIRE, P., FUSAR-POLI, P., ZANDI, M. S., LEWIS, G. and DAVID, A. S. (2020). Psychiatric and neuropsychiatric presentations associated with severe coronavirus infections: A systematic review and meta-analysis with comparison to the Covid-19 pandemic. *Lancet Psychiatry* **7** 611–627.

ROY, U. B., MANTEL, S., JACOBSON, M. and FERRIS, A. (2017). P2. 08-005 treating Cachexia-Anorexia in lung cancer patients: Understanding the patient perspective on novel treatment approaches: Topic: Patient's voice, patient's information. *J. Thorac. Oncol.* **12** S1112.

SKRONDAL, A. and RABE-HESKETH, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models. Interdisciplinary Statistics*. CRC Press/CRC, Boca Raton, FL. MR2059021 https://doi.org/10.1201/9780203489437

SKRONDAL, A. and RABE-HESKETH, S. (2007). Latent variable modelling: A survey. *Scand. J. Stat.* **34** 712–745. MR2396936 https://doi.org/10.1111/j.1467-9469.2007.00573.x

SOSA, J. and BETANCOURT, B. (2022). A latent space model for multilayer network data. *Comput. Statist. Data Anal.* **169** Paper No. 107432. MR4369145 https://doi.org/10.1016/j.csda.2022.107432

SPIEGEL, D. and GIESE-DAVIS, J. (2003). Depression and cancer: Mechanisms and disease progression. *Biol. Psychiatry* **54** 269–282. https://doi.org/10.1016/s0006-3223(03)00566-3

SWANSON, T., TEIXEIRA, A. S., RICHSON, B. N., YING, L., HILLS, T., FORBUSH, K. T., WATSON, D. and STELLA, M. (2021). Cognitive networks identify dimensions of distress in suicide notes: Anxiety, emotional profiles, and "words not said".

VAN DER EIJK, M., FABER, M. J., AARTS, J. W., KREMER, J. A., MUNNEKE, M., BLOEM, B. R. et al. (2013). Using online health communities to deliver patient-centered care to people with chronic conditions. *J. Med. Internet Res.* **15** e2476.

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

ZHANG, C. and ZHANG, J. (2010). InForCE: Forum data crawling with information extraction. In 2010 4*th International Universal Communication Symposium* 367–373. IEEE Press, New York.

Zhang, S., O'Carroll Bantum, E., Owen, J., Bakken, S. and Elhadad, N. (2017). Online cancer communities as informatics intervention for social support: Conceptualization, characterization, and impact. *J. Amer. Med. Inform. Assoc.* **24** 451–459.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 159–183
https://doi.org/10.1214/23-AOAS1784

GENERATIVE MACHINE LEARNING METHODS FOR MULTIVARIATE ENSEMBLE POSTPROCESSING

BY JIEYU CHEN[1,a], TIM JANKE[2,c], FLORIAN STEINKE[2,d] AND SEBASTIAN LERCH[1,b]

[1]*Institute of Economics, Karlsruhe Institute of Technology,* [a]*jieyu.chen@kit.edu,* [b]*sebastian.lerch@kit.edu*
[2]*Energy Information Networks & Systems, Technical University of Darmstadt,* [c]*tim.janke@gmx.net,*
[d]*florian.steinke@eins.tu-darmstadt.de*

Ensemble weather forecasts based on multiple runs of numerical weather prediction models typically show systematic errors and require postprocessing to obtain reliable forecasts. Accurately modeling multivariate dependencies is crucial in many practical applications, and various approaches to multivariate postprocessing have been proposed where ensemble predictions are first postprocessed separately in each margin and multivariate dependencies are then restored via copulas. These two-step methods share common key limitations, in particular, the difficulty to include additional predictors in modeling the dependencies. We propose a novel multivariate postprocessing method based on generative machine learning to address these challenges. In this new class of nonparametric data-driven distributional regression models, samples from the multivariate forecast distribution are directly obtained as output of a generative neural network. The generative model is trained by optimizing a proper scoring rule, which measures the discrepancy between the generated and observed data, conditional on exogenous input variables. Our method does not require parametric assumptions on univariate distributions or multivariate dependencies and allows for incorporating arbitrary predictors. In two case studies on multivariate temperature and wind speed forecasting at weather stations over Germany, our generative model shows significant improvements over state-of-the-art methods and particularly improves the representation of spatial dependencies.

REFERENCES

ADEL, T., GHAHRAMANI, Z. and WELLER, A. (2018). Discovering interpretable representations for both deep generative and discriminative models. In *Proceedings of the* 35*th International Conference on Machine Learning* **80** 50–59. PMLR, Atlanta, GA, USA.

ALEXANDER, C., COULON, M., HAN, Y. and MENG, X. (2022). Evaluating the discrimination ability of proper multi-variate scoring rules. *Ann. Oper. Res.* 1–27.

ALLEN, S., GINSBOURGER, D. and ZIEGEL, J. (2022). Evaluating forecasts for high-impact events using transformed kernel scores. Preprint. Available at arXiv:2202.12732.

BARAN, S. and LERCH, S. (2015). Log-normal distribution based ensemble model output statistics models for probabilistic wind-speed forecasting. *Q. J. R. Meteorol. Soc.* **141** 2289–2299.

BARAN, S. and LERCH, S. (2016). Mixture EMOS model for calibrating ensemble forecasts of wind speed. *Environmetrics* **27** 116–130. MR3481324 https://doi.org/10.1002/env.2380

BARAN, S. and MÖLLER, A. (2015). Joint probabilistic forecasting of wind speed and temperature using Bayesian model averaging. *Environmetrics* **26** 120–132. MR3324906 https://doi.org/10.1002/env.2316

BARAN, S., SZOKOL, P. and SZABÓ, M. (2021). Truncated generalized extreme value distribution-based ensemble model output statistics model for calibration of wind speed ensemble forecasts. *Environmetrics* **32** e2678. MR4309587 https://doi.org/10.1002/env.2678

BAUER, P., THORPE, A. and BRUNET, G. (2015). The quiet revolution of numerical weather prediction. *Nature* **525** 47–55. https://doi.org/10.1038/nature14956

Key words and phrases. Generative machine learning, ensemble postprocessing, multivariate postprocessing, probabilistic forecasting, weather forecasting.

BERGSTRA, J., YAMINS, D. and COX, D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *Proceedings of the* 30*th International Conference on Machine Learning* **28** 115–123. PMLR, Atlanta, GA, USA.

BOUALLEGUE, Z. B., HAIDEN, T., WEBER, N. J., HAMILL, T. M. and RICHARDSON, D. S. (2020). Accounting for representativeness in the verification of ensemble precipitation forecasts. *Mon. Weather Rev.* **148** 2049–2062.

BOUALLÈGUE, Z. B., HEPPELMANN, T., THEIS, S. E. and PINSON, P. (2016). Generation of scenarios from calibrated ensemble forecasts with a dual-ensemble copula-coupling approach. *Mon. Weather Rev.* **144** 4737–4750.

BOUGEAULT, P., TOTH, Z. et al. (2010). The THORPEX interactive grand global ensemble. *Bull. Am. Meteorol. Soc.* **91** 1059–1072.

BREMNES, J. B. (2020). Ensemble postprocessing using quantile function regression based on neural networks and Bernstein polynomials. *Mon. Weather Rev.* **148** 403–414.

CHALOULOS, G. and LYGEROS, J. (2007). Effect of wind correlation on aircraft conflict probability. *J. Guid. Control Dyn.* **30** 1742–1752.

CHAPMAN, W. E., MONACHE, L. D., ALESSANDRINI, S., SUBRAMANIAN, A. C., RALPH, F. M., XIE, S.-P., LERCH, S. and HAYATBINI, N. (2022). Probabilistic predictions from deterministic atmospheric river forecasts with deep learning. *Mon. Weather Rev.* **150** 215–234.

CHEN, J., JANKE, T., STEINKE, F. and LERCH, S. (2024). Supplement to "Generative machine learning methods for multivariate ensemble postprocessing." https://doi.org/10.1214/23-AOAS1784SUPPA, https://doi.org/10.1214/23-AOAS1784SUPPB

CHEN, X., DUAN, Y., HOUTHOOFT, R., SCHULMAN, J., SUTSKEVER, I. and ABBEEL, P. (2016). InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in Neural Information Processing Systems* **29**. Curran Associates, Red Hook.

CLARK, M., GANGOPADHYAY, S., HAY, L., RAJAGOPALAN, B. and WILBY, R. (2004). The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J. Hydrometeorol.* **5** 243–262.

CLEVERT, D.-A., UNTERTHINER, T. and HOCHREITER, S. (2015). Fast and accurate deep network learning by exponential linear units (ELUs). Preprint. Available at arXiv:1511.07289.

DAI, Y. and HEMRI, S. (2021). Spatially coherent postprocessing of cloud cover ensemble forecasts. *Mon. Weather Rev.* **149** 3923–3937.

DZIUGAITE, G. K., ROY, D. M. and GHAHRAMANI, Z. (2015). Training generative neural networks via Maximum Mean Discrepancy optimization. Preprint. Available at arXiv:1505.03906.

FELDMANN, K., SCHEUERER, M. and THORARINSDOTTIR, T. L. (2015). Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression. *Mon. Weather Rev.* **143** 955–971.

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

GNEITING, T. and RAFTERY, A. E. (2007a). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548 https://doi.org/10.1198/016214506000001437

GNEITING, T., RAFTERY, A. E., WESTVELD, A. H. and GOLDMAN, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Weather Rev.* **133** 1098–1118.

GOODFELLOW, I. J., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. and BENGIO, Y. (2014). Generative adversarial nets. In *NIPS'*14: *Proceedings of the* 27*th International Conference on Neural Information Processing Systems—Volume* 2 2672–2680. MIT Press, Cambridge, MA, USA.

GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716

GUI, J., SUN, Z., WEN, Y., TAO, D. and YE, J. (2021). A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Trans. Knowl. Data Eng.* 1–1.

HARRIS, L., MCRAE, A. T. T., CHANTRY, M., DUEBEN, P. D. and PALMER, T. N. (2022). A generative deep learning approach to stochastic downscaling of precipitation forecasts. *J. Adv. Model. Earth Syst.* **14** e2022MS003120.

HAUPT, S. E., CHAPMAN, W., ADAMS, S. V., KIRKWOOD, C., HOSKING, J. S., ROBINSON, N. H., LERCH, S. and SUBRAMANIAN, A. C. (2021). Towards implementing artificial intelligence post-processing in weather and climate: Proposed actions from the Oxford 2019 workshop. *Philos. Trans. R. Soc. A* **379** 20200091. MR4236151 https://doi.org/10.1098/rsta.2020.0091

HESS, P., DRÜKE, M., PETRI, S., STRNAD, F. and BOERS, N. (2022). Physically constrained generative adversarial networks for improving precipitation fields from Earth system models. *Nature Machine Intelligence* **4** 828–839. https://doi.org/10.1038/s42256-022-00540-1

HU, Y., SCHMEITS, M. J., VAN ANDEL, S. J., VERKADE, J. S., XU, M., SOLOMATINE, D. P. and LIANG, Z. (2016). A stratified sampling approach for improved sampling from a calibrated ensemble forecast distribution. *J. Hydrometeorol.* **17** 2405–2417.

JANKE, T., GHANMI, M. and STEINKE, F. (2021). Implicit generative copulas. In *Advances in Neural Information Processing Systems* **34** 26028–26039. Curran Associates, Red Hook.

JANKE, T. and STEINKE, F. (2020). Probabilistic multivariate electricity price forecasting using implicit generative ensemble post-processing. In 2020 *International Conference on Probabilistic Methods Applied to Power Systems* (*PMAPS*) 1–6. IEEE, Los Alamitos.

JORDAN, A., KRÜGER, F. and LERCH, S. (2019). Evaluating probabilistic forecasts with scoringRules. *J. Stat. Softw.* **90** 1–37.

KINGMA, D. P. and BA, J. (2014). Adam: A method for stochastic optimization. Preprint. Available at arXiv:1412.6980.

LAKATOS, M., LERCH, S., HEMRI, S. and BARAN, S. (2023). Comparison of multivariate post-processing methods using global ECMWF ensemble forecasts. *Q. J. R. Meteorol. Soc.* **149** 856–877.

LAKSHMINARAYANAN, B., PRITZEL, A. and BLUNDELL, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In *NIPS'17*: *Proceedings of the* 31*st International Conference on Neural Information Processing Systems* 6405–6416. Curran Associates Inc., Red Hook, NY, USA.

LANG, M. N., LERCH, S., MAYR, G. J., SIMON, T., STAUFFER, R. and ZEILEIS, A. (2020). Remember the past: A comparison of time-adaptive training schemes for non-homogeneous regression. *Nonlinear Process. Geophys.* **27** 23–34.

LANG, M. N., MAYR, G. J., STAUFFER, R. and ZEILEIS, A. (2019). Bivariate Gaussian models for wind vectors in a distributional regression framework. *Adv. Stat. Climatol. Meteorol. Oceanogr.* **5** 115–132.

LEINONEN, J., NERINI, D. and BERNE, A. (2021). Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Trans. Geosci. Remote Sens.* **59** 7211–7223.

LERCH, S. and BARAN, S. (2017). Similarity-based semilocal estimation of post-processing models. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 29–51. MR3611676 https://doi.org/10.1111/rssc.12153

LERCH, S., BARAN, S., MÖLLER, A., GROSS, J., SCHEFZIK, R., HEMRI, S. and GRAETER, M. (2020). Simulation-based comparison of multivariate ensemble post-processing methods. *Nonlinear Process. Geophys.* **27** 349–371.

LERCH, S. and POLSTERER, K. L. (2022). Convolutional autoencoders for spatially-informed ensemble post-processing. International Conference on Learning Representations (ICLR) 2022—AI for Earth and Space Science Workshop. https://doi.org/10.48550/arXiv.2204.05102

LERCH, S. and THORARINSDOTTIR, T. L. (2013). Comparison of non-homogeneous regression models for probabilistic wind speed forecasting. *Tellus, Ser. A Dyn. Meteorol. Oceanogr.* **65** 21206.

LERCH, S., THORARINSDOTTIR, T. L., RAVAZZOLO, F. and GNEITING, T. (2017). Forecaster's dilemma: Extreme events and forecast evaluation. *Statist. Sci.* **32** 106–127. MR3634309 https://doi.org/10.1214/16-STS588

LI, Y., SWERSKY, K. and ZEMEL, R. (2015). Generative moment matching networks. In *Proceedings of the* 32*nd International Conference on Machine Learning* **37** 1718–1727. PMLR, Lille, France.

MATHESON, J. E. and WINKLER, R. L. (1976). Scoring rules for continuous probability distributions. *Manage. Sci.* **22** 1087–1096.

MCGOVERN, A., LAGERQUIST, R., GAGNE, D. J., JERGENSEN, G. E., ELMORE, K. L., HOMEYER, C. R. and SMITH, T. (2019). Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Am. Meteorol. Soc.* **100** 2175–2199.

MOHAMED, S. and LAKSHMINARAYANAN, B. (2016). Learning in Implicit Generative Models. Preprint. Available at arXiv:1610.03483.

MÖLLER, A., LENKOSKI, A. and THORARINSDOTTIR, T. L. (2013). Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Q. J. R. Meteorol. Soc.* **139** 982–991.

MUSCHINSKI, T., LANG, M. N., MAYR, G. J., MESSNER, J. W., ZEILEIS, A. and SIMON, T. (2022). Predicting power ramps from joint distributions of future wind speeds. *Wind Energy Sci.* **7** 2393–2405.

NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2197664 https://doi.org/10.1007/s11229-005-3715-x

PACCHIARDI, L., ADEWOYIN, R., DUEBEN, P. and DUTTA, R. (2021). Probabilistic forecasting with generative networks via scoring rule minimization. Preprint. Available at arXiv:2112.08217.

PANTILLON, F., LERCH, S., KNIPPERTZ, P. and CORSMEIER, U. (2018). Forecasting wind gusts in winter storms using a calibrated convection-permitting ensemble. *Q. J. R. Meteorol. Soc.* **144** 1864–1881.

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A. et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12** 2825–2830. MR2854348

PENNINGTON, J., SOCHER, R. and MANNING, C. (2014). GloVe: Global vectors for word representation. In *Proceedings of the* 2014 *Conference on Empirical Methods in Natural Language Processing* (*EMNLP*) 1532–1543. Association for Computational Linguistics, Doha, Qatar.

PERRONE, E., SCHICKER, I. and LANG, M. N. (2020). A case study of empirical copula methods for the statistical correction of forecasts of the ALADIN-LAEF system. *Meteorol. Z.* **29** 277–288.

PETROPOULOS, F. et al. (2022). Forecasting: Theory and practice. *Int. J. Forecast.* **38** 705–871.

PINSON, P. and GIRARD, R. (2012). Evaluating the quality of scenarios of short-term wind power generation. *Appl. Energy* **96** 12–20.

PINSON, P. and MESSNER, J. W. (2018). Chapter 9—Application of postprocessing for renewable energy. In *Statistical Postprocessing of Ensemble Forecasts* 241–266. Elsevier, Amsterdam.

PRICE, I. and RASP, S. (2022). Increasing the accuracy and resolution of precipitation forecasts using deep generative models. In *Proceedings of the* 25*th International Conference on Artificial Intelligence and Statistics* **151** 10555–10571. PMLR, Atlanta, GA, USA.

RASP, S. and LERCH, S. (2018). Neural networks for postprocessing ensemble weather forecasts. *Mon. Weather Rev.* **146** 3885–3900.

RAVURI, S., LENC, K., WILLSON, M. et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature* **597** 672–677.

SCHEFZIK, R. (2016). A similarity-based implementation of the Schaake shuffle. *Mon. Weather Rev.* **144** 1909–1921.

SCHEFZIK, R. and MÖLLER, A. (2018). Chapter 4—Ensemble postprocessing methods incorporating dependence structures. In *Statistical Postprocessing of Ensemble Forecasts* 91–125. Elsevier, Amsterdam.

SCHEFZIK, R., THORARINSDOTTIR, T. L. and GNEITING, T. (2013). Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statist. Sci.* **28** 616–640. MR3161590 https://doi.org/10.1214/13-STS443

SCHEUERER, M. and HAMILL, T. M. (2015). Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Mon. Weather Rev.* **143** 1321–1334.

SCHEUERER, M., HAMILL, T. M., WHITIN, B., HE, M. and HENKEL, A. (2017). A method for preferential selection of dates in the Schaake shuffle approach to constructing spatiotemporal forecast fields of temperature and precipitation. *Water Resour. Res.* **53** 3029–3046.

SCHEUERER, M. and MÖLLER, D. (2015). Probabilistic wind speed forecasting on a grid based on ensemble model output statistics. *Ann. Appl. Stat.* **9** 1328–1349. MR3418725 https://doi.org/10.1214/15-AOAS843

SCHEUERER, M., SWITANEK, M. B., WORSNOP, R. P. and HAMILL, T. M. (2020). Using artificial neural networks for generating probabilistic subseasonal precipitation forecasts over California. *Mon. Weather Rev.* **148** 3489–3506.

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

SCHULZ, B. and LERCH, S. (2022a). Machine learning methods for postprocessing ensemble forecasts of wind gusts: A systematic comparison. *Mon. Weather Rev.* **150** 235–257.

SCHULZ, B. and LERCH, S. (2022b). Aggregating distribution forecasts from deep ensembles. Preprint. Available at arXiv:2204.02291.

SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866 https://doi.org/10.1214/13-AOS1140

SKLAR, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **8** 229–231. MR0125600

SLOUGHTER, M., GNEITING, T. and RAFTERY, A. E. (2010). Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Amer. Statist. Assoc.* **105** 25–35. MR2757189 https://doi.org/10.1198/jasa.2009.ap08615

SONG, Y. and ERMON, S. (2020). Improved techniques for training score-based generative models. In *NIPS*'20: *Proceedings of the* 34*th International Conference on Neural Information Processing Systems* **1043** 12438–12448. Curran Associates Inc., Red Hook, NY, USA.

SZÉKELY, G. J. and RIZZO, M. L. (2013). Energy statistics: A class of statistics based on distances. *J. Stat. Plan. Inference.* **143** 1249–1272.

TAILLARDAT, M., MESTRE, O., ZAMO, M. and NAVEAU, P. (2016). Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. *Mon. Weather Rev.* **144** 2375–2393.

THORARINSDOTTIR, T. L. and GNEITING, T. (2010). Probabilistic forecasts of wind speed: Ensemble model ouput statistics by using heteroscedastic censored regression. *J. Roy. Statist. Soc. Ser. A* **173** 371–388. MR2751882 https://doi.org/10.1111/j.1467-985X.2009.00616.x

VANNITSEM, S., BREMNES, J. B., DEMAEYER, J., EVANS, G. R., FLOWERDEW, J., HEMRI, S., LERCH, S., ROBERTS, N., THEIS, S. et al. (2021). Statistical postprocessing for weather forecasts: Review, challenges, and avenues in a big data world. *Bull. Am. Meteorol. Soc.* **102** E681–E699.

VAN SCHAEYBROECK, B. and VANNITSEM, S. (2015). Ensemble post-processing using member-by-member approaches: Theoretical aspects. *Q. J. R. Meteorol. Soc.* **141** 807–818.

WILKS, D. S. (2015). Multivariate ensemble model output statistics using empirical copulas. *Q. J. R. Meteorol. Soc.* **141** 945–952.

WORSNOP, R. P., SCHEUERER, M., HAMILL, T. M. and LUNDQUIST, J. K. (2018). Generating wind power scenarios for probabilistic ramp event prediction using multivariate statistical post-processing. *Wind Energy Sci.* **3** 371–393.

ZHOU, B. (2022). Interpreting generative adversarial networks for interactive image generation. In *XxAI—Beyond Explainable AI: International Workshop, Held in Conjunction with ICML* 2020, *July* 18, 2020, *Vienna, Austria, Revised and Extended Papers* 167–175. Springer, Cham.

ZIEL, F. and BERK, K. (2019). Multivariate forecasting evaluation: On sensitive and strictly proper scoring rules. Preprint. Available at arXiv:1910.07325.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 184–198
https://doi.org/10.1214/23-AOAS1785

APPLIED REGRESSION ANALYSIS OF CORRELATIONS FOR CORRELATED DATA

BY JIE HU[1,a], YU CHEN[1,b], CHENLEI LENG[2,c] AND CHENG YONG TANG[3,d]

[1]*International Institute of Finance, School of Management, University of Science and Technology of China,* [a]*hujie@mail.ustc.edu.cn,* [b]*cyu@ustc.edu.cn*

[2]*Department of Statistics, University of Warwick,* [c]*C.Leng@warwick.ac.uk*

[3]*Department of Statistics, Operations, and Data Science, Temple University,* [d]*yongtang@temple.edu*

Correlated data are ubiquitous in today's data-driven society. While regression models for analyzing means and variances of responses of interest are relatively well developed, the development of these models for analyzing the correlations is largely confined to longitudinal data, a special form of sequentially correlated data. This paper proposes a new method for the analysis of correlations to fully exploit the use of covariates for general correlated data. In a renewed analysis of the classroom data, a highly unbalanced multilevel clustered data with within-class and within-school correlations, our method reveals informative insights on these structures not previously known. In another analysis of the malaria immune response data in Benin, a longitudinal study with time-dependent covariates where the exact times of the observations are not available, our approach again provides promising new results. At the heart of our approach is a new generalized z-transformation that converts correlation matrices, constrained to be positive definite, to vectors with unrestricted support and is order-invariant. These two properties enable us to develop regression analysis incorporating covariates for the modelling of correlations via the use of maximum likelihood.

REFERENCES

ADJAKOSSA, E. H., SADISSOU, I., HOUNKONNOU, M. N. and NUEL, G. (2016). Multivariate longitudinal analysis with bivariate correlation test. *PLoS ONE* **11** e0159649. https://doi.org/10.1371/journal.pone.0159649

ARCHAKOV, I. and HANSEN, P. R. (2021). A new parametrization of correlation matrices. *Econometrica* **89** 1699–1715. MR4325198 https://doi.org/10.3982/ecta16910

BAAYEN, R. H., DAVIDSON, D. J. and BATES, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* **59** 390–412.

BARR, D. J., LEVY, R., SCHEEPERS, C. and TILY, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68** 255–278.

BATES, D. (2006). lmer, p-values and all that.

BATES, D., KLIEGL, R., VASISHTH, S. and BAAYEN, H. (2015). Parsimonious mixed models. arXiv preprint. Available at arXiv:1506.04967.

BOLKER, B. et al. (2020). GLMM FAQ.

CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Stat.* **25** 573–578. MR0065087 https://doi.org/10.1214/aoms/1177728725

FARAWAY, J. J. (2016). *Extending the Linear Model with R*: *Generalized Linear, Mixed Effects and Nonparametric Regression Models*. *Chapman & Hall/CRC Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. Second edition [of MR2192856]. MR3617626

FITZMAURICE, G., DAVIDIAN, M., VERBEKE, G. and MOLENBERGHS, G., eds. (2009). *Longitudinal Data Analysis*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR1500110

HILL, H. C., ROWAN, B. and BALL, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *Am. Educ. Res. J.* **42** 371–406.

HU, J., CHEN, Y., LENG, C. and TANG, C. Y. (2024). Supplement to "Applied regression analysis of correlations for correlated data." https://doi.org/10.1214/23-AOAS1785SUPP

Key words and phrases. Correlogram, correlated data analysis, correlation matrix, generalized z-transformation, regression modeling, testing correlation structures.

LAIRD, N. M. and WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38** 963–974.

LENG, C., ZHANG, W. and PAN, J. (2010). Semiparametric mean-covariance regression analysis for longitudinal data. *J. Amer. Statist. Assoc.* **105** 181–193. With supplementary material available online. MR2656048 https://doi.org/10.1198/jasa.2009.tm08485

LUKE, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* **49** 1494–1502.

MANDEL, F., GHOSH, R. P. and BARNETT, I. (2023). Neural networks for clustered and longitudinal data using mixed effects models. *Biometrics* **79** 711–721. MR4606309 https://doi.org/10.1111/biom.13615

MÜLLER, S., SCEALY, J. L. and WELSH, A. H. (2013). Model selection in linear mixed models. *Statist. Sci.* **28** 135–167. MR3112403 https://doi.org/10.1214/12-sts410

PAN, J. and MACKENZIE, G. (2003). On modelling mean-covariance structures in longitudinal studies. *Biometrika* **90** 239–244. MR1966564 https://doi.org/10.1093/biomet/90.1.239

POURAHMADI, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86** 677–690. MR1723786 https://doi.org/10.1093/biomet/86.3.677

POURAHMADI, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* **87** 425–435. MR1782488 https://doi.org/10.1093/biomet/87.2.425

WEST, B. T., WELCH, K. B. and GALECKI, A. T. (2006). *Linear Mixed Models: A Practical Guide Using Statistical Software*. CRC Press, Boca Raton.

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. MR0640163 https://doi.org/10.2307/1912526

YE, H. and PAN, J. (2006). Modelling of covariance structures in generalised estimating equations for longitudinal data. *Biometrika* **93** 927–941. MR2285080 https://doi.org/10.1093/biomet/93.4.927

ZHANG, W. and LENG, C. (2012). A moving average Cholesky factor model in covariance modelling for longitudinal data. *Biometrika* **99** 141–150. MR2899669 https://doi.org/10.1093/biomet/asr068

ZHANG, W., LENG, C. and TANG, C. Y. (2015). A joint modelling approach for longitudinal studies. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 219–238. MR3299406 https://doi.org/10.1111/rssb.12065

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 199–223
https://doi.org/10.1214/23-AOAS1786

A BEHAVIORAL APPROACH TO REPEATED BAYESIAN SECURITY GAMES

BY WILLIAM CABALLERO[1,a], JAKE COOLEY[1,b], DAVID BANKS[2,c] AND PHILLIP JENKINS[3,d]

[1]*Department of Mathematical Sciences, United States Air Force Academy,* [a]*william.caballero@afacademy.af.edu,* [b]*jake.cooley.2@us.af.mil*

[2]*Department of Statistical Science, Duke University,* [c]*david.banks@duke.edu*

[3]*Department of Operational Sciences, Air Force Institute of Technology,* [d]*phillip.jenkins.4@us.af.mil*

The prevalence of security threats to organizational defense demands models that support real-world policymaking. Security games are a potent tool in this regard; however, although canonical models effectively allocate limited resources, they generally do not consider adaptive, boundedly rational adversaries. Empirical findings suggest this characterization describes real-world human behavior, so the development of decision-support frameworks against such adversaries is a critical need. We examine a family of policies applicable to repeated games in which a boundedly rational adversary is modeled using a behavioral-economic theory of learning, that is, experience-weighted attraction learning. These policies take into account realistic uncertainty about the competition by adopting the perspective of adversarial risk analysis. Using Bayesian reasoning, these repeated games are decomposed into multiarm bandit problems. A collection of cost-function approximation policies are given to solve these problems. The efficacy of our approach is shown via extensive computational testing on a defense-related case study.

REFERENCES

ALBRECHT, S. V. and STONE, P. (2018). Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence* **258** 66–95. MR3771158 https://doi.org/10.1016/j.artint.2018.01.002

AN, B., ORDÓÑEZ, F., TAMBE, M., SHIEH, E., YANG, R., BALDWIN, C., DIRENZO, J. III, MORETTI, K., MAULE, B. et al. (2013). A deployed quantal response-based patrol planning system for the us coast guard. *Interfaces* **43** 400–420.

AUER, P., CESA-BIANCHI, N., FREUND, Y. and SCHAPIRE, R. E. (1995). Gambling in a rigged casino: The adversarial multi-armed bandit problem. In 36*th Annual Symposium on Foundations of Computer Science* (*Milwaukee*, *WI*, 1995) 322–331. IEEE Comput. Soc., Los Alamitos, CA. MR1619094 https://doi.org/10.1109/SFCS.1995.492488

AUER, P., CESA-BIANCHI, N., FREUND, Y. and SCHAPIRE, R. E. (2002). The nonstochastic multiarmed bandit problem. *SIAM J. Comput*. **32** 48–77. MR1954855 https://doi.org/10.1137/S0097539701398375

BANKS, D., GALLEGO, V., NAVEIRO, R. and RÍOS INSUA, D. (2022). Adversarial risk analysis: An overview. *Wiley Interdiscip. Rev.*: *Comput. Stat*. **14** Paper No. e1530, 16. MR4376781 https://doi.org/10.1002/wics.1530

BANKS, D., PETRALIA, F. and WANG, S. (2011). Adversarial risk analysis: Borel games. *Appl. Stoch. Models Bus. Ind*. **27** 72–86. MR2828655 https://doi.org/10.1002/asmb.890

BANKS, D. L., RIOS, J. and RÍOS INSUA, D. (2016). *Adversarial Risk Analysis*. CRC Press, Boca Raton, FL. MR3409217

BARBERIS, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *J. Econ. Perspect*. **27** 173–96.

BAZARAA, M. S., SHERALI, H. D. and SHETTY, C. M. (2013). *Nonlinear Programming*: *Theory and Algorithms*. Wiley, New York.

BERGER, J. O. (2013). *Statistical Decision Theory and Bayesian Analysis*, Springer, Berlin.

CABALLERO, W. N., BANKS, D. and WU, K. (2022). Defense and security planning under resource uncertainty and multi-period commitments. *Naval Res. Logist*. **69** 1009–1026. MR4478228 https://doi.org/10.1002/nav.22071

Key words and phrases. Adversarial risk analysis, behavioral game theory, multiarm bandits, repeated games.

CABALLERO, W. N. and LUNDAY, B. J. (2023). *Military and Security Applications*: *Behavioral Modeling*. Springer, Cham (pp. 1–14). https://doi.org/10.1007/978-3-030-54621-2_759-1

CABALLERO, W. N., LUNDAY, B. J. and UBER, R. P. (2021). Identifying behaviorally robust strategies for normal form games under varying forms of uncertainty. *European J. Oper. Res.* **288** 971–982. MR4152747 https://doi.org/10.1016/j.ejor.2020.06.022

CAMERER, C. and HO, T.-H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica* **67** 827–874.

CAMERER, C. and HO, T. H. (1998). Experience-weighted attraction learning in coordination games: Probability rules, heterogeneity, and time-variation. *J. Math. Psych.* **42** 305–326. https://doi.org/10.1006/jmps.1998.1217

CAMERER, C. F. (2004). Behavioral game theory: Predicting human behavior in strategic situations. *Adv. Behav. Econ.* 374–392.

CAMERER, C. F., HO, T.-H. and CHONG, J.-K. (2002). Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *J. Econom. Theory* **104** 137–188.

CANO, J., RÍOS INSUA, D., TEDESCHI, A. and TURHAN, U. (2016). Security economics: An adversarial risk analysis approach to airport protection. *Ann. Oper. Res.* **245** 359–378. MR3552277 https://doi.org/10.1007/s10479-014-1690-7

CEN, S., CHENG, C., CHEN, Y., WEI, Y. and CHI, Y. (2022). Fast global convergence of natural policy gradient methods with entropy regularization. *Oper. Res.* **70** 2563–2578. MR4484422 https://doi.org/10.1287/opre.2021.2151

COX, J. C., SHACHAT, J. and WALKER, M. (2001). An experiment to evaluate Bayesian learning of Nash equilibrium play. *Games Econom. Behav.* **34** 11–33.

EJAZ, M., JOE, S. and JOSHI, C. (2021). Adversarial risk analysis for auctions using mirror equilibrium and Bayes Nash equilibrium. *Decis. Anal.* **18** 185–202. MR4328802 https://doi.org/10.1287/deca.2021.0425

EREV, I. and HARUVY, E. (2013). Learning and the economics of small decisions. In *The Handbook of Experimental Economics* **2** 1–123.

EREV, I. and RAPOPORT, A. (1998). Coordination, "magic," and reinforcement learning in a market entry game. *Games Econom. Behav.* **23** 146–175.

FRAZIER, P. I. (2018). A tutorial on Bayesian optimization. ArXiv preprint. Available at arXiv:1807.02811.

GALLEGO, V., NAVEIRO, R., INSUA, D. R. and OTEIZA, D. G.-U. (2019). Opponent aware reinforcement learning. ArXiv preprint. Available at arXiv:1908.08773.

GARIVIER, A. and MOULINES, E. (2011). On upper-confidence bound policies for switching bandit problems. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **6925** 174–188. Springer, Heidelberg. MR2889051 https://doi.org/10.1007/978-3-642-24412-4_16

HO, T.-H., PARK, S.-E. and SU, X. (2021). A Bayesian level-k model in n-person games. *Manage. Sci.* **67** 1622–1638.

HO, T. H., CAMERER, C. F. and CHONG, J.-K. (2007). Self-tuning experience weighted attraction learning in games. *J. Econom. Theory* **133** 177–198. MR2386761 https://doi.org/10.1016/j.jet.2005.12.008

JENKINS, P. R., ROBBINS, M. J. and LUNDAY, B. J. (2021). Approximate dynamic programming for military medical evacuation dispatching policies. *INFORMS J. Comput.* **33** 2–26. MR4214592 https://doi.org/10.1287/ijoc.2019.0930

JOSEPH, V. R. (2016). Space-filling designs for computer experiments: A review. *Qual. Eng.* **28** 28–35.

JOSHI, C., RIOS, J. and INSUA, D. R. (2020). Insider threat modeling: An adversarial risk analysis approach. *IEEE Trans. Inform. Forensics Secur.*

KAHNEMAN, D., SLOVIC, S. P., SLOVIC, P. and TVERSKY, A. (1982). *Judgment Under Uncertainty*: *Heuristics and Biases*. Cambridge Univ. Press, Cambridge.

KAR, D., FANG, F., DELLE FAVE, F. M., SINTOV, N. and TAMBE, M. (2015). "A game of thrones" when human behavior models compete in repeated Stackelberg security games. In *Proceedings of the* 2015 *International Conference on Autonomous Agents and Multiagent Systems*. 1381–1390.

KAR, D., FORD, B., GHOLAMI, S., FANG, F., PLUMPTRE, A., TAMBE, M., DRICIRU, M., WANYAMA, F., RWETSIBA, A. et al. (2017). Cloudy with a chance of poaching: Adversary behavior modeling and forecasting with real-world poaching data. In *Proceedings of the* 2016 *International Conference on Autonomous Agents and Multiagent Systems* ACM, New York.

KIEKINTVELD, C., JAIN, M., TSAI, J., PITA, J., ORDÓNEZ, F. and TAMBE, M. (2009). Computing optimal randomized resource allocations for massive security games. In *Proceedings of the* 8th *International Conference on Autonomous Agents and Multiagent Systems-Volume* 1 689–696.

KIM, B. and TEWARI, A. (2019). On the optimality of perturbations in stochastic and adversarial multi-armed bandit problems. ArXiv preprint. Available at arXiv:1902.00610.

KOLLER, D. and MILCH, B. (2003). Multi-agent influence diagrams for representing and solving games. *Games Econom. Behav.* **45** 181–221. First World Congress of the Game Theory Society (Bilbao, 2000). MR2022867 https://doi.org/10.1016/S0899-8256(02)00544-4

KORZHYK, D., YIN, Z., KIEKINTVELD, C., CONITZER, V. and TAMBE, M. (2011). Stackelberg vs. Nash in security games: An extended investigation of interchangeability, equivalence, and uniqueness. *J. Artificial Intelligence Res.* **41** 297–327. MR2863313 https://doi.org/10.1613/jair.3269

LATTIMORE, T. and SZEPESVÁRI, C. (2020). *Bandit Algorithms*. Cambridge Univ. Press, Cambridge.

MEI, J., XIAO, C., SZEPESVARI, C. and SCHUURMANS, D. (2020). On the global convergence rates of softmax policy gradient methods. In *International Conference on Machine Learning* 6820–6829. PMLR, Online.

MERSEREAU, A. J., RUSMEVICHIENTONG, P. and TSITSIKLIS, J. N. (2009). A structured multiarmed bandit problem and the greedy policy. *IEEE Trans. Automat. Control* **54** 2787–2802. MR2583719 https://doi.org/10.1109/TAC.2009.2031725

MYERS, R. H., MONTGOMERY, D. C. and ANDERSON-COOK, C. M. (2016). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 4th ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3497873

NAVEIRO, R., REDONDO, A., RÍOS INSUA, D. and RUGGERI, F. (2019). Adversarial classification: An adversarial risk analysis approach. *Internat. J. Approx. Reason.* **113** 133–148. MR3983223 https://doi.org/10.1016/j.ijar.2019.07.003

NEU, G. and BARTÓK, G. (2013). An efficient algorithm for learning with semi-bandit feedback. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **8139** 234–248. Springer, Heidelberg. MR3133069 https://doi.org/10.1007/978-3-642-40935-6_17

NGUYEN, T., YANG, R., AZARIA, A., KRAUS, S. and TAMBE, M. (2013). Analyzing the effectiveness of adversary modeling in security games. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

NGUYEN, T. H., WANG, Y., SINHA, A. and WELLMAN, M. P. (2019). Deception in finitely repeated security games. In *Proceedings of the AAAI Conference on Artificial Intelligence* 2133–2140.

PITA, J., TAMBE, M., KIEKINTVELD, C., CULLEN, S. and STEIGERWALD, E. (2011). Guards: Game theoretic security allocation on a national scale. In *The* 10*th International Conference on Autonomous Agents and Multiagent Systems-Volume* 1 37–44.

POWELL, W. B. (2007). *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. *Wiley Series in Probability and Statistics*. Wiley-Interscience, Hoboken, NJ. MR2347698 https://doi.org/10.1002/9780470182963

POWELL, W. B. (2009). What you should know about approximate dynamic programming. *Naval Res. Logist.* **56** 239–249. MR2503223 https://doi.org/10.1002/nav.20347

POWELL, W. B. (2016). Perspectives of approximate dynamic programming. *Ann. Oper. Res.* **241** 319–356. MR3509421 https://doi.org/10.1007/s10479-012-1077-6

POWELL, W. B. (2019). A unified framework for stochastic optimization. *European J. Oper. Res.* **275** 795–821. MR3913407 https://doi.org/10.1016/j.ejor.2018.07.014

POWELL, W. B. and RYZHOV, I. O. (2012). *Optimal Learning*. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2917881 https://doi.org/10.1002/9781118309858

PUTTA, S. R. and AGRAWAL, S. (2022). Scale-free adversarial multi armed bandits. *Proc. Mach. Learn. Res.* (*PMLR*) **167** 910–930. Algorithmic Learning Theory 2022.

RIOS INSUA, D., RIOS, J. and BANKS, D. (2009). Adversarial risk analysis. *J. Amer. Statist. Assoc.* **104** 841–854. MR2541597 https://doi.org/10.1198/jasa.2009.0155

RIOS, J. and INSUA, D. R. (2012). Adversarial risk analysis for counterterrorism modeling. *Risk Anal.: Int. J.* **32** 894–915.

RIOS INSUA, D., BANKS, D. and RÍOS, J. (2016). Modeling opponents in adversarial risk analysis. *Risk Anal.*, **36** 742–755.

RIOS INSUA, D., COUCE-VIEIRA, A., RUBIO, J. A., PIETERS, W., LABUNETS, K. and RASINES, D. (2021). An adversarial risk analysis framework for cybersecurity. *Risk Anal.* **41** 16–36.

RYZHOV, I. O., POWELL, W. B. and FRAZIER, P. I. (2012). The knowledge gradient algorithm for a general class of online learning problems. *Oper. Res.* **60** 180–195. MR2911666 https://doi.org/10.1287/opre.1110.0999

SESSA, P. G., BOGUNOVIC, I., KAMGARPOUR, M. and KRAUSE, A. (2020). Learning to play sequential games versus unknown opponents. ArXiv preprint. Available at arXiv:2007.05271.

SIEBERT, J., VON WINTERFELDT, D. and JOHN, R. S. (2016). Identifying and structuring the objectives of the Islamic State of Iraq and the Levant (ISIL) and its followers. *Decis. Anal.* **13** 26–50. MR3488805 https://doi.org/10.1287/deca.2015.0324

SINHA, A., FANG, F., AN, B., KIEKINTVELD, C. and TAMBE, M. (2018). Stackelberg security games: Looking beyond a decade of success. In *IJCAI*.

SLIVKINS, A. et al. (2020). Introduction to multi-armed bandits. *Found. Trends Mach. Learn.* **12** 1–286.

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

TAMBE, M., JIANG, A. X., AN, B., JAIN, M. et al. (2014). Computational game theory for security: Progress and challenges. In *AAAI Spring Symposium on Applied Computational Game Theory*.

TVERSKY, A. and KAHNEMAN, D. (1971). Belief in the law of small numbers. *Psychol. Bull.* **76** 105.

TVERSKY, A. and KOEHLER, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychol. Rev.* **101** 547–567.

WANG, J., CLARK, S. C., LIU, E. and FRAZIER, P. I. (2020). Parallel Bayesian global optimization of expensive functions. *Oper. Res.* **68** 1850–1865. MR4217270 https://doi.org/10.1287/opre.2019.1966

WANG, W., DI MAIO, F. and ZIO, E. (2019). Adversarial risk analysis to allocate optimal defense resources for protecting cyber–physical systems from cyber attacks. *Risk Anal.* **39** 2766–2785.

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

WITLIN, L. (2008). Of note: Mirror-imaging and its dangers. *SAIS Rev. Int. Aff.* **28** 89–90.

XU, H., TRAN-THANH, L. and JENNINGS, N. R. (2016). Playing repeated security games with no prior knowledge. In *AAMAS* 104–112.

XU, L., BONDI, E., FANG, F., PERRAULT, A., WANG, K. and TAMBE, M. (2021). Dual-mandate patrols: Multi-armed bandits for green security. In *Proceedings of the AAAI Conference on Artificial Intelligence* 14974–14982.

YANG, R., KIEKINTVELD, C., ORDÓÑEZ, F., TAMBE, M. and JOHN, R. (2013). Improving resource allocation strategies against human adversaries in security games: An extended study. *Artificial Intelligence* **195** 440–469. MR3024211 https://doi.org/10.1016/j.artint.2012.11.004

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 224–245
https://doi.org/10.1214/23-AOAS1787

NONNEGATIVE TENSOR COMPLETION FOR DYNAMIC COUNTERFACTUAL PREDICTION ON COVID-19 PANDEMIC

BY YAOMING ZHEN[1,a] AND JUNHUI WANG[2,b]

[1]*School of Data Science, City University of Hong Kong,* [a]*yzhen8-c@my.cityu.edu.hk*
[2]*Dapartment of Statistics, The Chinese University of Hong Kong,* [b]*junhuiwang@cuhk.edu.hk*

The COVID-19 pandemic has been a worldwide health crisis for the past three years, casting unprecedented challenges for policymakers in different countries and regions. While one country or region can only implement one social mobility restriction policy at a given time, it is of great interest for policy makers to decide whether to elevate or deelevate the restriction policy from time to time. This article proposes a novel nonnegative tensor completion method to predict the potential counterfactual outcomes of multifaceted social mobility restriction policies over time. The proposed method builds upon a low-rank tensor decomposition of the pandemic data, which also explicitly characterizes the ordinal nature of the mobility restriction strength and the smooth trend of the pandemic evolution over time. Its application to the COVID-19 pandemic data reveals some interesting facts regarding the impact of social mobility restriction policy on the spread of the virus. The effectiveness of the proposed method is also supported by its asymptotic estimation consistency and extensive numerical experiments on the synthetic datasets.

REFERENCES

AGARWAL, A., SHAH, D. and SHEN, D. (2022). Synthetic interventions. ArXiv preprint. Available at arXiv:2006.07691.

ATHEY, S., BAYATI, M., DOUDCHENKO, N., IMBENS, G. and KHOSRAVI, K. (2021). Matrix completion methods for causal panel data models. *J. Amer. Statist. Assoc.* **116** 1716–1730. MR4353709 https://doi.org/10.1080/01621459.2021.1891924

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

BAI, J. and NG, S. (2021). Matrix completion, counterfactuals, and factor analysis of missing data. *J. Amer. Statist. Assoc.* **116** 1746–1763. MR4353711 https://doi.org/10.1080/01621459.2021.1967163

BI, X., QU, A. and SHEN, X. (2018). Multilayer tensor factorization with applications to recommender systems. *Ann. Statist.* **46** 3308–3333. MR3852653 https://doi.org/10.1214/17-AOS1659

CAI, C., LI, G., POOR, H. V. and CHEN, Y. (2019). Nonconvex low-rank tensor completion from noisy data. *Adv. Neural Inf. Process. Syst.* **32**.

CAI, J.-F., LI, J. and XIA, D. (2022a). Generalized low-rank plus sparse tensor estimation by fast Riemannian optimization. *J. Amer. Statist. Assoc.* 1–17.

CAI, J.-F., LI, J. and XIA, D. (2022b). Provable tensor-train format tensor completion by Riemannian optimization. *J. Mach. Learn. Res.* **23** Paper No. [123], 77. MR4577075

CHEN, B., SUN, T., ZHOU, Z. and ZENG, Y. (2019). Nonnegative tensor completion via low-rank Tucker decomposition: Model and algorithm. *IEEE Access* 95903–95914.

DONG, E., DU, H. and GARDNER, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20** 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1

FAN, J., MASINI, R. and MEDEIROS, M. C. (2022). Do we exploit all information for counterfactual analysis? Benefits of factor models and idiosyncratic correction. *J. Amer. Statist. Assoc.* **117** 574–590. MR4436297 https://doi.org/10.1080/01621459.2021.2004895

FAN, Y., LU, X., ZHAO, J., FU, H. and LIU, Y. (2022). Estimating individualized treatment rules for treatments with hierarchical structure. *Electron. J. Stat.* **16** 737–784. MR4366820 https://doi.org/10.1214/21-ejs1948

Key words and phrases. Causal inference, imputation, informative missing, latent factor, tensor decomposition.

FOGARTY, C. B. (2020). Studentized sensitivity analysis for the sample average treatment effect in paired observational studies. *J. Amer. Statist. Assoc.* **115** 1518–1530. MR4143482 https://doi.org/10.1080/01621459.2019.1632072

GOUTTE, C. and AMINI, M.-R. (2010). Probabilistic tensor factorization and model selection. In *Tensors, Kernels, and Machine Learning* (*TKLM* 2010) 1–4.

HÖFLER, M. (2005). Causal inference based on counterfactuals. *BMC Med. Res. Methodol.* **5** 1–12.

HU, J., LEE, C. and WANG, M. (2022). Generalized tensor decomposition with features on multiple modes. *J. Comput. Graph. Statist.* **31** 204–218. MR4387221 https://doi.org/10.1080/10618600.2021.1978471

JAIN, P. and OH, S. (2014). Provable tensor factorization with missing data. *Adv. Neural Inf. Process. Syst.* **27**.

JI, P. and JIN, J. (2016). Coauthorship and citation networks for statisticians. *Ann. Appl. Stat.* **10** 1779–1812. MR3592033 https://doi.org/10.1214/15-AOAS896

JOHANSSON, F., SHALIT, U. and SONTAQ, D. (2016). Learning representation for counterfactual inference. In *International Conference on Machine Learning* 3020–3029.

KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056 https://doi.org/10.1137/07070111X

LAUER, S. A., GRANTZ, K. H., BI, Q., JONES, F. K., ZHENG, Q., MEREDITH, H. R., AZMAN, A. S., REICH, N. G. and LESSLER, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **172** 577–582. https://doi.org/10.7326/M20-0504

LIU, A. and MOITRA, A. (2020). Tensor completion made practical. In *Advances in Neural Information Processing Systems* **33** 18905–8916.

MANDAL, D. and PARKES, D. (2022). Weighted tensor completion for time-series causal inference. ArXiv preprint. Available at arXiv:1902.04646.

MCALOON, C., COLLINS, Á., HUNT, K., BARBER, A., BYRNE, A. W., CASEY, M., GRIFFIN, J., LANE, E., MCEVOY, D. et al. (2020). Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. *BMJ Open* **10** e039652.

MEN, K., LI, Y., WANG, X., ZHANG, G., HU, J., GAO, Y., HAN, A. and LIU, W. (2023). Estimate the incubation period of coronavirus 2019 (COVID-19). *Comput. Biol. Med.* **158** 106794.

MO, W. and LIU, Y. (2022). Efficient learning of optimal individualized treatment rules for heteroscedastic or misspecified treatment-free effect models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 440–472. MR4412993 https://doi.org/10.1111/rssb.12474

MO, W., QI, Z. and LIU, Y. (2021). Learning optimal distributionally robust individualized treatment rules. *J. Amer. Statist. Assoc.* **116** 659–674. MR4270012 https://doi.org/10.1080/01621459.2020.1796359

NEWEY, W. K. and STOULI, S. (2022). Heterogeneous coefficients, control variables and identification of multiple treatment effects. *Biometrika* **109** 865–872. MR4472853 https://doi.org/10.1093/biomet/asab060

PAPADOGEORGOU, G., IMAI, K., LYALL, J. and LI, F. (2022). Causal inference with spatio-temporal data: Estimating the effects of airstrikes on insurgent violence in Iraq. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1969–1999. MR4515563 https://doi.org/10.1111/rssb.12548

PEARL, J. (2009). Causal inference in statistics: An overview. *Stat. Surv.* **3** 96–146. MR2545291 https://doi.org/10.1214/09-SS057

POULOS, J., ALBANESE, A., MERCATANTI, A. and LI, F. (2022). Retrospective causal inference via matrix completion, with an evaluation of the effect of European integration on cross-border employment. ArXiv preprint. Available at arXiv:2106.00788.

QI, Z., LIU, D., FU, H. and LIU, Y. (2020). Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various outcomes. *J. Amer. Statist. Assoc.* **115** 678–691. MR4107672 https://doi.org/10.1080/01621459.2018.1529597

QUICK, C., DEY, R. and LIN, X. (2021). Regression models for understanding COVID-19 epidemic dynamics with incomplete data. *J. Amer. Statist. Assoc.* **116** 1561–1577. MR4353694 https://doi.org/10.1080/01621459.2021.2001339

ROHE, K., QIN, T. and YU, B. (2016). Co-clustering directed graphs to discover asymmetries and directional communities. *Proc. Natl. Acad. Sci. USA* **113** 12679–12684. MR3576189 https://doi.org/10.1073/pnas.1525793113

ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

SINHA, S. and CHAKRABORTY, M. (2022). Causal analysis and prediction of human mobility in the US during the COVID-19 pandemic. ArXiv preprint. Available at arXiv:2111.12272.

SUN, W. W. and LI, L. (2019). Dynamic tensor clustering. *J. Amer. Statist. Assoc.* **114** 1894–1907. MR4047308 https://doi.org/10.1080/01621459.2018.1527701

WILSON, N., KVALSVIG, A., BARNARD, L. T. and BAKER, M. G. (2020). Case-fatality risk estimates for COVID-19 calculated by using a lag time for fatality. *Emerg. Infec. Dis.* **26** 1339–1441. https://doi.org/10.3201/eid2606.200320

XIA, D. and YUAN, M. (2019). On polynomial time methods for exact low-rank tensor completion. *Found. Comput. Math.* **19** 1265–1313. MR4029842 https://doi.org/10.1007/s10208-018-09408-6

XU, Y. and YIN, W. (2013). A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM J. Imaging Sci.* **6** 1758–1789. MR3105787 https://doi.org/10.1137/120887795

XUE, F. and QU, A. (2021). Integrating multisource block-wise missing data in model selection. *J. Amer. Statist. Assoc.* **116** 1914–1927. MR4353722 https://doi.org/10.1080/01621459.2020.1751176

YADLOWSKY, S., PELLEGRINI, F., LIONETTO, F., BRAUNE, S. and TIAN, L. (2021). Estimation and validation of ratio-based conditional average treatment effects using observational data. *J. Amer. Statist. Assoc.* **116** 335–352. MR4227698 https://doi.org/10.1080/01621459.2020.1772080

YAN, H., ZHU, Y., GU, J., HUANG, Y., SUN, H., ZHANG, X., WANG, Y., QIU, Y. and CHEN, S. X. (2021). Better strategies for containing COVID-19 pandemic: A study of 25 countries via a vSIADR model. *Proc. R. Soc. A* **477** Paper No. 20200440, 25. MR4258333

YE, Y., ZHANG, Q., CAO, Z., CHEN, F. Y. YAN, H., STANLEY, H. E. and ZENG, D. D. (2021). Impact of export restrictions on the global personal protective equipment trade network during COVID-19. *Adv. Theory Simul.* 2100352.

YUAN, M. and ZHANG, C.-H. (2016). On tensor completion via nuclear norm minimization. *Found. Comput. Math.* **16** 1031–1068. MR3529132 https://doi.org/10.1007/s10208-015-9269-5

YUAN, M. and ZHANG, C.-H. (2017). Incoherent tensor norms and their applications in higher order tensor completion. *IEEE Trans. Inf. Theory* **63** 6753–6766. MR3707566 https://doi.org/10.1109/TIT.2017.2724549

ZHANG, J., SUN, W. W. and LI, L. (2020). Mixed-effect time-varying network model and application in brain connectivity analysis. *J. Amer. Statist. Assoc.* **115** 2022–2036. MR4189774 https://doi.org/10.1080/01621459.2019.1677242

ZHANG, X. and NG, M. K. (2022). Sparse nonnegative tensor factorization and completion with noisy observations. *IEEE Trans. Inf. Theory* **68** 2551–2572. MR4413569

ZHANG, Y., BI, X., TANG, N. and QU, A. (2021). Dynamic tensor recommender systems. *J. Mach. Learn. Res.* **22** Paper No. 65, 35. MR4253758

ZHEN, Y. and WANG, J. (2024). Supplement to "Nonnegative tensor completion for dynamic counterfactual prediction on COVID-19 pandemic." https://doi.org/10.1214/23-AOAS1787SUPP

ZHOU, J., SUN, W. W., ZHANG, J. and LI, L. (2023). Partially observed dynamic tensor response regression. *J. Amer. Statist. Assoc.* **118** 424–439. MR4571132 https://doi.org/10.1080/01621459.2021.1938082

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 246–265
https://doi.org/10.1214/23-AOAS1788

DYNAMIC RISK PREDICTION FOR CERVICAL PRECANCER SCREENING WITH CONTINUOUS AND BINARY LONGITUDINAL BIOMARKERS

BY SIDDHARTH ROY[1,a], ANINDYA ROY[2,d], MEGAN A. CLARKE[3,e],
ANA GRADISSIMO[4,g], ROBERT D. BURK[5,h], NICOLAS WENTZENSEN[3,f],
PAUL S. ALBERT[1,b] AND DANPING LIU[1,c]

[1]*Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute,* [a]*siddharth.roy@nih.gov,* [b]*albertp@mail.nih.gov,* [c]*danping.liu@nih.gov*

[2]*Department of Mathematics and Statistics, University of Maryland Baltimore County,* [d]*anindya@umbc.edu*

[3]*Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute,* [e]*megan.clarke@nih.gov,* [f]*wentzenn@mail.nih.gov*

[4]*Department of Immunology, Memorial Sloan Kettering Cancer Center,* [g]*gradisa@mskcc.org*

[5]*Department of Pediatrics, Albert Einstein College of Medicine,* [h]*robert.burk@einsteinmed.org*

Dynamic risk prediction that incorporates longitudinal measurements of biomarkers is useful in identifying high-risk patients for better clinical management. Our work is motivated by the prediction of cervical precancers. Currently, Pap cytology is used to identify HPV-positive (HPV+) women at high-risk of cervical precancer, but cytology lacks accuracy and reproducibility. Molecular markers, like HPV DNA methylation, that are closely linked to the carcinogenic process show promise of improved risk stratification. We are interested in developing a dynamic risk model that uses all longitudinal biomarker information to improve precancer risk estimation. We propose a joint model to link both the continuous methylation biomarker and a binary cytology biomarker to the time to precancer outcome using shared random effects. The model uses a discretization of the time scale to allow for closed-form likelihood expressions, thereby avoiding potential high dimensional integration of the random effects. The method handles an interval-censored time-to-event outcome, due to intermittent clinical visits, incorporates sampling weights to deal with stratified sampling data and can provide immediate and five-year risk estimates that may inform clinical decision-making. Applying the method to longitudinally measured HPV methylation data improves risk stratification for triage of HPV+ women.

REFERENCES

ALBERT, P. S. (2019). Shared random parameter models: A legacy of the biostatistics program at the national heart, lung and blood institute. *Stat. Med.* **38** 501–511. MR3902594 https://doi.org/10.1002/sim.8011

ARNOLD, B. C. (2009). Flexible univariate and multivariate models based on hidden truncation. *J. Statist. Plann. Inference* **139** 3741–3749. MR2553758 https://doi.org/10.1016/j.jspi.2009.05.013

BAART, S. J., BOERSMA, E. and RIZOPOULOS, D. (2019). Joint models for longitudinal and time-to-event data in a case-cohort design. *Stat. Med.* **38** 2269–2281. MR3939816 https://doi.org/10.1002/sim.8113

BARRETT, J., DIGGLE, P., HENDERSON, R. and TAYLOR-ROBINSON, D. (2015). Joint modelling of repeated measurements and time-to-event outcomes: Flexible model specification and exact likelihood inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 131–148. MR3299402 https://doi.org/10.1111/rssb.12060

BARRETT, J. and SU, L. (2017). Dynamic predictions using flexible joint models of longitudinal and time-to-event data. *Stat. Med.* **36** 1447–1460. MR3631971 https://doi.org/10.1002/sim.7209

BROWN, E. R., IBRAHIM, J. G. and DEGRUTTOLA, V. (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics* **61** 64–73. MR2129202 https://doi.org/10.1111/j.0006-341X.2005.030929.x

---

Key words and phrases. Cervical precancer, dynamic risk prediction, joint model, longitudinal biomarkers, shared random effects.

CHEUNG, L. C., EGEMEN, D., CHEN, X., KATKI, H. A., DEMARCO, M., WISER, A. L., PERKINS, R. B., GUIDO, R. S., WENTZENSEN, N. and SCHIFFMAN, M. (2020). 2019 ASCCP risk-based management consensus guidelines: Methods for risk estimation, recommended management, and validation. *J. Low. Genit. Tract Dis.* **24** 90.

CHEUNG, L. C., PAN, Q., HYUN, N., SCHIFFMAN, M., FETTERMAN, B., CASTLE, P. E., LOREY, T. and KATKI, H. A. (2017). Mixture models for undiagnosed prevalent disease and interval-censored incident disease: Applications to a cohort assembled from electronic health records. *Stat. Med.* **36** 3583–3595. MR3696511 https://doi.org/10.1002/sim.7380

CLARKE, M. A., GRADISSIMO, A., SCHIFFMAN, M., LAM, J., SOLLECITO, C. C., FETTERMAN, B., LOREY, T., POITRAS, N., RAINE-BENNETT, T. R., CASTLE, P. E. et al. (2018). Human papillomavirus DNA methylation as a biomarker for cervical precancer: Consistency across 12 genotypes and potential impact on management of HPV-positive women. *Clin. Cancer Res.* **24** 2194–2202.

CURRY, S. J., KRIST, A. H., OWENS, D. K., BARRY, M. J., CAUGHEY, A. B., DAVIDSON, K. W., DOUBENI, C. A., EPLING, J. W., KEMPER, A. R., KUBIK, M. et al. (2018). Screening for cervical cancer: US preventive services task force recommendation statement. *Jama* **320** 674–686.

EGEMEN, D., CHEUNG, L. C., CHEN, X., DEMARCO, M., PERKINS, R. B., KINNEY, W., POITRAS, N., BEFANO, B., LOCKE, A., GUIDO, R. S. et al. (2020). Risk estimates supporting the 2019 ASCCP risk-based management consensus guidelines. *J. Low. Genit. Tract Dis.* **24** 132.

FAUCETT, C. L. and THOMAS, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Stat. Med.* **15** 1663–1685.

FONTHAM, E. T. H., WOLF, A. M. D., CHURCH, T. R. ETZIONI, R., FLOWERS, C. R., HERZIG, A., GUERRA, C. E., OEFFINGER, K. C., SHIH, Y.-C. T., WALTER, L. C. et al. (2020). Cervical cancer screening for individuals at average risk: 2020 guideline update from the American cancer society. *CA Cancer J. Clin.* **70** 321–346.

GAGE, J. C., SCHIFFMAN, M., KATKI, H. A., CASTLE, P. E., FETTERMAN, B., WENTZENSEN, N., POITRAS, N. E., LOREY, T., CHEUNG, L. C. and KINNEY, W. K. (2014). Reassurance against future risk of precancer and cancer conferred by a negative human papillomavirus test. *J. Natl. Cancer Inst.* **106**. https://doi.org/10.1093/jnci/dju153

GENZ, A. (1993). Comparison of methods for the computation of multivariate normal probabilities. *Computing Science and Statistics* 400–400.

GENZ, A. and BRETZ, F. (2009). *Computation of Multivariate Normal and t Probabilities. Lecture Notes in Statistics* **195**. Springer, Dordrecht. MR2840595 https://doi.org/10.1007/978-3-642-01689-9

GENZ, A., BRETZ, F., MIWA, T., MI, X., LEISCH, F., SCHEIPL, F. and HOTHORN, T. (2020). mvtnorm: Multivariate Normal and t Distributions R package version 1.1-1.

HYUN, N., CHEUNG, L. C., PAN, Q., SCHIFFMAN, M. and KATKI, H. A. (2017). Flexible risk prediction models for left or interval-censored data from electronic health records. *Ann. Appl. Stat.* **11** 1063–1084. MR3693558 https://doi.org/10.1214/17-AOAS1036

MAUFF, K., STEYERBERG, E., KARDYS, I., BOERSMA, E. and RIZOPOULOS, D. (2020). Joint models with multiple longitudinal outcomes and a time-to-event outcome: A corrected two-stage approach. *Stat. Comput.* **30** 999–1014. MR4108688 https://doi.org/10.1007/s11222-020-09927-9

PAKMAN, A. and PANINSKI, L. (2014). Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *J. Comput. Graph. Statist.* **23** 518–542. MR3215823 https://doi.org/10.1080/10618600.2013.788448

PAPAGEORGIOU, G., MAUFF, K., TOMER, A. and RIZOPOULOS, D. (2019). An overview of joint modeling of time-to-event and longitudinal outcomes. *Annu. Rev. Stat. Appl.* **6** 223–240. MR3939519 https://doi.org/10.1146/annurev-statistics-030718-105048

PERKINS, R. B., GUIDO, R. S., CASTLE, P. E., CHELMOW, D., EINSTEIN, M. H., GARCIA, F., HUH, W. K., KIM, J. J., MOSCICKI, A.-B., NAYAR, R. et al. (2020). 2019 ASCCP risk-based management consensus guidelines for abnormal cervical cancer screening tests and cancer precursors. *J. Low. Genit. Tract Dis.* **24** 102.

RIZOPOULOS, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC press, Boca Raton.

RIZOPOULOS, D. and GHOSH, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Stat. Med.* **30** 1366–1380. MR2828959 https://doi.org/10.1002/sim.4205

RIZOPOULOS, D., TAYLOR, J. M. G., VAN ROSMALEN, J., STEYERBERG, E. W. and TAKKENBERG, J. J. M. (2016). Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. *Biostatistics* **17** 149–164. MR3449857 https://doi.org/10.1093/biostatistics/kxv031

ROY, S., ROY, A., CLARKE, M. A., GRADISSIMO, A., BURK, R. D., WENTZENSEN, N., ALBERT, P. S. and LIU, D. (2024). Supplement to "Dynamic risk prediction for cervical precancer screening with continuous and binary longitudinal biomarkers." https://doi.org/10.1214/23-AOAS1788SUPPA, https://doi.org/10.1214/23-AOAS1788SUPPB

SCHIFFMAN, M., HYUN, N., RAINE-BENNETT, T. R., KATKI, H., FETTERMAN, B., GAGE, J. C., CHEUNG, L. C., BEFANO, B., POITRAS, N., LOREY, T. et al. (2016). A cohort study of cervical screening using partial HPV typing and cytology triage. *Int. J. Cancer* **139** 2606–2615.

SOLOMON, D., DAVEY, D., KURMAN, R., MORIARTY, A., O'CONNOR, D. PREY, M. RAAB, S. SHERMAN, M. WILBUR, D. WRIGHT JR, T. et al. (2002). The 2001 Bethesda system: Terminology for reporting results of cervical cytology. *Jama* **287** 2114–2119.

TOMER, A., NIEBOER, D., ROOBOL, M. J., STEYERBERG, E. W. and RIZOPOULOS, D. (2019). Personalized schedules for surveillance of low-risk prostate cancer patients. *Biometrics* **75** 153–162. MR3953716 https://doi.org/10.1111/biom.12940

TSIATIS, A. A. and DAVIDIAN, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statist. Sinica* **14** 809–834. MR2087974

WENTZENSEN, N., ARBYN, M., BERKHOF, J., BOWER, M., CANFELL, K., EINSTEIN, M., FARLEY, C., MONSONEGO, J. and FRANCESCHI, S. (2017). Eurogin 2016 roadmap: How HPV knowledge is changing screening practice. *Int. J. Cancer* **140** 2192–2200. https://doi.org/10.1002/ijc.30579

WU, Y. and COOK, R. J. (2022). Assessing the accuracy of predictive models with interval-censored data. *Biostatistics* **23** 18–33. MR4366033 https://doi.org/10.1093/biostatistics/kxaa011

WULFSOHN, M. S. and TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53** 330–339. MR1450186 https://doi.org/10.2307/2533118

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 266–302
https://doi.org/10.1214/23-AOAS1789

BAYESIAN MODEL-BASED CLUSTERING FOR POPULATIONS OF NETWORK DATA

BY ANASTASIA MANTZIOU[1,a], SIMÓN LUNAGÓMEZ[2,b] AND ROBIN MITRA[3,c]

[1]*Department of Finance and Economics, The Alan Turing Institute,* [a]*amantziou@turing.ac.uk*
[2]*Departamento de Estadística, ITAM,* [b]*simon.lunagomez@itam.mx*
[3]*Department of Statistical Science, University College London,* [c]*robin.mitra@ucl.ac.uk*

There is increasing appetite for analysing populations of network data due to the fast-growing body of applications demanding such methods. While methods exist to provide readily interpretable summaries of heterogeneous network populations, these are often descriptive or ad hoc, lacking any formal justification. In contrast, principled analysis methods often provide results difficult to relate back to the applied problem of interest. Motivated by two complementary applied examples, we develop a Bayesian framework to appropriately model complex heterogeneous network populations, while also allowing analysts to gain insights from the data and make inferences most relevant to their needs. The first application involves a study in computer science measuring human movements across a university. The second analyses data from neuroscience investigating relationships between different regions of the brain. While both applications entail analysis of a heterogeneous population of networks, network sizes vary considerably. We focus on the problem of clustering the elements of a network population, where each cluster is characterised by a network representative. We take advantage of the Bayesian machinery to simultaneously infer the cluster membership, the representatives, and the community structure of the representatives, thus allowing intuitive inferences to be made. The implementation of our method on the human movement study reveals interesting movement patterns of individuals in clusters, readily characterised by their network representative. For the brain networks application, our model reveals a cluster of individuals with different network properties of particular interest in neuroscience. The performance of our method is additionally validated in extensive simulation studies.

REFERENCES

AIROLDI, E. M. and BLOCKER, A. W. (2013). Estimating latent processes on a network from indirect measurements. *J. Amer. Statist. Assoc.* **108** 149–164. MR3174609 https://doi.org/10.1080/01621459.2012.756328

ARROYO, J., ATHREYA, A., CAPE, J., CHEN, G., PRIEBE, C. E. and VOGELSTEIN, J. T. (2021). Inference for multiple heterogeneous networks with a common invariant subspace. *J. Mach. Learn. Res.* **22** Paper No. 142, 49. MR4318498

ARROYO RELIÓN, J. D., KESSLER, D., LEVINA, E. and TAYLOR, S. F. (2019). Network classification with applications to brain connectomics. *Ann. Appl. Stat.* **13** 1648–1677. MR4019153 https://doi.org/10.1214/19-AOAS1252

BALACHANDRAN, P., KOLACZYK, E. D. and VILES, W. D. (2017). On the propagation of low-rate measurement error to subgraph counts in large networks. *J. Mach. Learn. Res.* **18** Paper No. 61, 33. MR3687604

BASSETT, D. S. and BULLMORE, E. D. (2006). Small-world brain networks. *Neuroscientist* **12** 512–523.

BUDIAJI, W. (2019). kmed: Distance-based K-medoids R package version 0.3.0.

CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2009). Handling sparsity via the horseshoe. In *Artificial Intelligence and Statistics*. 73–80. PMLR, U.S.

CHANG, J., KOLACZYK, E. D. and YAO, Q. (2022). Estimation of subgraph densities in noisy networks. *J. Amer. Statist. Assoc.* **117** 361–374. MR4399091 https://doi.org/10.1080/01621459.2020.1778482

Key words and phrases. Bayesian models, clustering, mixture models, populations of network data, object data analysis.

CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604 https://doi.org/10.1214/14-AOS1272

CRADDOCK, R. C., JAMES, G. A., HOLTZHEIMER, P. E. III, HU, X. and MAYBERG, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* **33** 1914–1928.

DIQUIGIOVANNI, J. and SCARPA, B. (2019). Analysis of association football playing styles: An innovative method to cluster networks. *Stat. Model.* **19** 28–54. MR3903741 https://doi.org/10.1177/1471082X18808628

DONNAT, C. and HOLMES, S. (2018). Tracking network dynamics: A survey using graph distances. *Ann. Appl. Stat.* **12** 971–1012. MR3834292 https://doi.org/10.1214/18-AOAS1176

DURANTE, D., DUNSON, D. B. and VOGELSTEIN, J. T. (2017). Nonparametric Bayes modeling of populations of networks. *J. Amer. Statist. Assoc.* **112** 1516–1530. MR3750873 https://doi.org/10.1080/01621459.2016.1219260

FIELDS, S. and SONG, O. (1989). A novel genetic system to detect protein–protein interactions. *Nature* **340** 245–246.

FRÜHWIRTH-SCHNATTER, S. and MALSINER-WALLI, G. (2019). From here to infinity: Sparse finite versus Dirichlet process mixtures in model-based clustering. *Adv. Data Anal. Classif.* **13** 33–64. MR3935190 https://doi.org/10.1007/s11634-018-0329-y

GINESTET, C. E., LI, J., BALACHANDRAN, P., ROSENBERG, S. and KOLACZYK, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.* **11** 725–750. MR3693544 https://doi.org/10.1214/16-AOAS1015

GOLLINI, I. and MURPHY, T. B. (2016). Joint modeling of multiple network views. *J. Comput. Graph. Statist.* **25** 246–265. MR3474046 https://doi.org/10.1080/10618600.2014.978006

HANDCOCK, M. S. and GILE, K. J. (2010). Modeling social networks from sampled data. *Ann. Appl. Stat.* **4** 5–25. MR2758082 https://doi.org/10.1214/08-AOAS221

HEITJAN, D. F. and RUBIN, D. B. (1990). Inference from coarse data via multiple imputation with application to age heaping. *J. Amer. Statist. Assoc.* **85** 304–314.

HEITJAN, D. F. and RUBIN, D. B. (1991). Ignorability and coarse data. *Ann. Statist.* **19** 2244–2253. MR1135174 https://doi.org/10.1214/aos/1176348396

HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262 https://doi.org/10.1198/016214502388618906

INRA and LEGER, J.-B. (2015). blockmodels: Latent and stochastic block model estimation by a 'V–EM' algorithm R package version 1.1.1.

JIANG, X., GOLD, D. and KOLACZYK, E. D. (2011). Network-based auto-probit modeling for protein function prediction. *Biometrics* **67** 958–966. MR2829270 https://doi.org/10.1111/j.1541-0420.2010.01519.x

JOSEPHS, N., LI, W. and KOLACZYK, E. D. (2021). Network recovery from unlabeled noisy samples. In 2021 55*th Asilomar Conference on Signals*, *Systems*, *and Computers* 1268–1273. IEEE, U.S.

KERIBIN, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā Ser. A* **62** 49–66. MR1769735

KIM, H. and PARK, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* **23** 1495–1502.

KIM, J. K. and HONG, M. (2012). Imputation for statistical inference with coarse data. *Canad. J. Statist.* **40** 604–618. MR2968401 https://doi.org/10.1002/cjs.11142

KOLACZYK, E. D., LIN, L., ROSENBERG, S., WALTERS, J. and XU, J. (2020). Averages of unlabeled networks: Geometric characterization and asymptotic behavior. *Ann. Statist.* **48** 514–538. MR4065172 https://doi.org/10.1214/19-AOS1820

KOSKINEN, J. H. ROBINS, G. L. WANG, P. PATTISON, P. E. (2013). Bayesian analysis for partially observed network data, missing ties, attributes and actors. *Soc. Netw.* **35** 514–527.

LE, C. M., LEVIN, K. and LEVINA, E. (2018). Estimating a network from multiple noisy realizations. *Electron. J. Stat.* **12** 4697–4740. MR3894068 https://doi.org/10.1214/18-ejs1521

LE, C. M. and LI, T. (2022). Linear regression and its inference on noisy network-linked data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1851–1885. MR4515560

LEVIN, K., ATHREYA, A., TANG, M., LYZINSKI, V., PARK, Y. and PRIEBE, C. E. (2017). A central limit theorem for an omnibus embedding of multiple random graphs and implications for multiscale network inference. ArXiv preprint. Available at arXiv:1705.09355.

LI, W., SUSSMAN, D. L. and KOLACZYK, E. D. (2021). Causal inference under network interference with noise. ArXiv preprint. Available at arXiv:2105.04518.

LIAO, X., VASILAKOS, A. V. and HE, Y. (2017). Small-world human brain networks: Perspectives and challenges. *Neurosci. Biobehav. Rev.* **77** 286–300.

LU, X. and MARRON, J. S. (2014). Analysis of juggling data: Object oriented data analysis of clustering in acceleration functions. *Electron. J. Stat.* **8** 1842–1847. MR3273603 https://doi.org/10.1214/14-EJS937D

LUNAGÓMEZ, S., OLHEDE, S. C. and WOLFE, P. J. (2021). Modeling network populations via graph distances. *J. Amer. Statist. Assoc.* **116** 2023–2040. MR4353730 https://doi.org/10.1080/01621459.2020.1763803

MALSINER-WALLI, G., FRÜHWIRTH-SCHNATTER, S. and GRÜN, B. (2016). Model-based clustering based on sparse finite Gaussian mixtures. *Stat. Comput.* **26** 303–324. MR3439375 https://doi.org/10.1007/s11222-014-9500-2

MANTZIOU, A., LUNAGÓMEZ, S. and MITRA, R. (2024). Supplement to "Bayesian model-based clustering for populations of network data." https://doi.org/10.1214/23-AOAS1789SUPPA, https://doi.org/10.1214/23-AOAS1789SUPPB

MARCHETTE, D. J. and HOHMAN, E. L. (2015). Utilizing covariates in partially observed networks. In 2015 18*th International Conference on Information Fusion* (*Fusion*) 166–172. IEEE, U.S.

MUKHERJEE, S. S., SARKAR, P. and LIN, L. (2017). On clustering network-valued data. In *Advances in Neural Information Processing Systems*.

NEAL, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Statist.* **9** 249–265. MR1823804 https://doi.org/10.2307/1390653

NEWMAN, M. E. J. (2018). Estimating network structure from unreliable measurements. *Phys. Rev. E* **98** 062321.

NIELSEN, A. M. and WITTEN, D. (2018). The multiple random dot product graph model. ArXiv preprint. Available at arXiv:1811.12172.

PEIXOTO, T. P. (2018). Reconstructing networks with unknown and heterogeneous errors. *Phys. Rev. X* **8** 041011.

PRASAD, G., JOSHI, S. H., NIR, T. M., TOGA, A. W., THOMPSON, P. M., ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE (ADNI) et al. (2015). Brain connectivity and novel network measures for Alzheimer's disease classification. *Neurobiol. Aging* **36** S121–S131.

PRIEBE, C. E., SUSSMAN, D. L., TANG, M. and VOGELSTEIN, J. T. (2015). Statistical inference on errorfully observed graphs. *J. Comput. Graph. Statist.* **24** 930–953. MR3432923 https://doi.org/10.1080/10618600.2014.951049

RASTELLI, R. and FRIEL, N. (2018). Optimal Bayesian estimators for latent variable cluster models. *Stat. Comput.* **28** 1169–1186. MR3850389 https://doi.org/10.1007/s11222-017-9786-y

RICHARDSON, S. and GREEN, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B* **59** 731–792. MR1483213 https://doi.org/10.1111/1467-9868.00095

RICHIARDI, J., ERYILMAZ, H., SCHWARTZ, S., VUILLEUMIER, P. and VAN DE VILLE, D. (2011). Decoding brain states from fMRI connectivity graphs. *NeuroImage* **56** 616–626.

SCHÜTZE, H., MANNING, C. D. and RAGHAVAN, P. (2008). *Introduction to Information Retrieval* 39. Cambridge Univ. Press, Cambridge.

SHAW, P., MIKUSZ, M., NURMI, P. and DAVIES, N. (2018). Tacita: A privacy preserving public display personalisation service. In *Proceedings of the* 2018 *ACM International Joint Conference and* 2018 *International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* 448–451.

SHEN, W., WANG, Y., BAI, X., WANG, H. and LATECKI, L.J. (2013). Shape clustering: Common structure discovery. *Pattern Recognit.* **46** 539–550.

SIGNORELLI, M. and WIT, E. C. (2020). Model-based clustering for populations of networks. *Stat. Model.* **20** 9–29. MR4052400 https://doi.org/10.1177/1471082X19871128

SONG, J. J., LEE, H.-J., MORRIS, J. S. and KANG, S. (2007). Clustering of time-course gene expression data using functional data analysis. *Comput. Biol. Chem.* **31** 265–274.

WANG, S., ARROYO, J., VOGELSTEIN, J. T. and PRIEBE, C. E. (2019). Joint embedding of graphs. *IEEE Trans. Pattern Anal. Mach. Intell.*

WHITE, J. G., SOUTHGATE, E., THOMSON, N. J. and BRENNER, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **314** 1–340.

YOUNG, J.-G., CANTWELL, G. T. and NEWMAN, M. E. J. (2020). Bayesian inference of network structure from unreliable data. *J. Complex Netw.* **8** cnaa046, 26. MR4228651 https://doi.org/10.1093/comnet/cnaa046

YOUNG, S. J. and SCHEINERMAN, E. R. (2007). Random dot product graph models for social networks. In *Algorithms and Models for the Web-Graph. Lecture Notes in Computer Science* **4863** 138–149. Springer, Berlin. MR2504912 https://doi.org/10.1007/978-3-540-77004-6_11

ZHANG, J., CHENG, W., WANG, Z., ZHANG, Z., LU, W., LU, G. and FENG, J. (2012). Pattern classification of large-scale functional brain networks: Identification of informative neuroimaging markers for epilepsy. *PLoS ONE* **7** e36733.

ZHAO, Y., WU, Y.-J., LEVINA, E. and ZHU, J. (2017). Link prediction for partially observed networks. *J. Comput. Graph. Statist.* **26** 725–733. MR3698680 https://doi.org/10.1080/10618600.2017.1286243

ZUO, X.-N., ANDERSON, J. S., BELLEC, P., BIRN, R. M., BISWAL, B. B., BLAUTZIK, J., BREITNER, J. C. S., BUCKNER, R. L., CALHOUN, V. D., CASTELLANOS, F. X. et al. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci. Data* **1** 1–13.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 303–327
https://doi.org/10.1214/23-AOAS1790

INFERRING CHANGES TO THE GLOBAL CARBON CYCLE WITH WOMBAT v2.0, A HIERARCHICAL FLUX-INVERSION FRAMEWORK

BY MICHAEL BERTOLACCI[1,a], ANDREW ZAMMIT-MANGION[1,b], ANDREW SCHUH[2,e], BEATA BUKOSA[3,f], JENNY A. FISHER[4,g], YI CAO[1,c], ALEYA KAUSHIK[5,h] AND NOEL CRESSIE[1,d]

[1]*School of Mathematics and Applied Statistics, University of Wollongong,* [a]*michael_bertolacci@uow.edu.au,* [b]*azm@uow.edu.au,* [c]*ycao@uow.edu.au,* [d]*ncressie@uow.edu.au*

[2]*Cooperative Institute for Research in the Atmosphere (CIRA), Colorado State University,* [e]*aschuh@atmos.colostate.edu*

[3]*National Institute of Water and Atmospheric Research (NIWA),* [f]*beata.bukosa@niwa.co.nz*

[4]*Centre for Atmospheric Chemistry, School of Earth, Atmospheric and Life Sciences, University of Wollongong,* [g]*jennyf@uow.edu.au*

[5]*Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado,* [h]*aleya.kaushik@noaa.gov*

The natural cycles of the surface-to-atmosphere fluxes of carbon dioxide (CO_2) and other important greenhouse gases are changing in response to human influences. These changes need to be quantified to understand climate change and its impacts, but this is difficult to do because natural fluxes occur over large spatial and temporal scales and cannot be directly observed. Flux inversion is a technique that estimates the spatiotemporal distribution of a gas' fluxes using observations of the gas' mole fraction and a chemical transport model. To infer trends in fluxes and identify phase shifts and amplitude changes in flux seasonal cycles, we construct a flux-inversion system that uses a novel spatially-varying time-series decomposition of the fluxes. We incorporate this decomposition into the Wollongong Methodology for Bayesian Assimilation of Trace-gases (WOMBAT, Zammit-Mangion et al., *Geosci. Model Dev.*, 15, 2022), a Bayesian hierarchical flux-inversion framework that yields posterior distributions for all unknowns in the underlying model. We also extend WOMBAT to accommodate physical constraints on the fluxes and to take direct in situ and flask measurements of trace-gas mole fractions as observations. We apply the new method, which we call WOMBAT v2.0, to a mix of satellite observations of CO_2 mole fraction from the Orbiting Carbon Observatory-2 (OCO-2) satellite and direct measurements of CO_2 mole fraction from a variety of sources. We estimate the changes in the natural cycles of CO_2 fluxes that occurred from January 2015 to December 2020, and compare our posterior estimates to those from an alternative method based on a bottom-up understanding of the physical processes involved. We find substantial trends in the fluxes, including that tropical ecosystems trended from being a net source to a net sink of CO_2 over the study period. We also find that the amplitude of the global seasonal cycle of ecosystem CO_2 fluxes increased over the study period by 0.11 PgC/month (an increase of 8%) and that the seasonal cycle of ecosystem CO_2 fluxes in the northern temperate and northern boreal regions shifted earlier in the year by 0.4–0.7 and 0.4–0.9 days, respectively (2.5th to 97.5th posterior percentiles), consistent with expectations for the carbon cycle under a warming climate.

REFERENCES

BARICHIVICH, J., BRIFFA, K. R., MYNENI, R. B., OSBORN, T. J., MELVIN, T. M., CIAIS, P., PIAO, S. and TUCKER, C. (2013). Large-scale variations in the vegetation growing season and annual cycle of atmospheric CO_2 at high northern latitudes from 1950 to 2011. *Glob. Change Biol.* **19** 3167–3183.

BASU, S., BAKER, D. F., CHEVALLIER, F., PATRA, P. K., LIU, J. and MILLER, J. B. (2018). The impact of transport model differences on CO_2 surface flux estimates from OCO-2 retrievals of column average CO_2. *Atmos. Chem. Phys.* **18** 7189–7215.

Key words and phrases. Carbon cycle, hierarchical, Bayesian, flux inversion, carbon dioxide.

BASU, S., GUERLET, S., BUTZ, A., HOUWELING, S., HASEKAMP, O., ABEN, I., KRUMMEL, P., STEELE, P., LANGENFELDS, R. et al. (2013). Global CO_2 fluxes estimated from GOSAT retrievals of total column CO_2. *Atmos. Chem. Phys.* **13** 8695–8717.

BERTOLACCI, M., ZAMMIT-MANGION, A., SCHUH, A., BUKOSA, B., FISHER, J., CAO, Y., KAUSHIK, A. and CRESSIE, N. (2024a). Supplement to “Inferring changes to the global carbon cycle with WOMBAT v2.0, a hierarchical flux-inversion framework.” https://doi.org/10.1214/23-AOAS1790SUPPA

BERTOLACCI, M., ZAMMIT-MANGION, A., SCHUH, A., BUKOSA, B., FISHER, J., CAO, Y., KAUSHIK, A. and CRESSIE, N. (2024b). wombat-v2-workflow: WOMBAT: A fully Bayesian global flux-inversion framework, version 2.0. https://doi.org/10.1214/23-AOAS1790SUPPB

BERTOLACCI, M., ZAMMIT-MANGION, A., SCHUH, A., BUKOSA, B., FISHER, J., CAO, Y., KAUSHIK, A. and CRESSIE, N. (2024c). WOMBAT v2.0 estimates of global GPP, respiration, and air-sea fluxes. https://doi.org/10.1214/23-AOAS1790SUPPC

BEY, I., JACOB, D. J., YANTOSCA, R. M., LOGAN, J. A., FIELD, B. D., FIORE, A. M., LI, Q., LIU, H. Y., MICKLEY, L. J. et al. (2001). Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *J. Geophys. Res., Atmos.* **106** 23073–23095.

BRUNSDON, C., FOTHERINGHAM, A. S. and CHARLTON, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geogr. Anal.* **28** 281–298.

BUITENWERF, R., ROSE, L. and HIGGINS, S. I. (2015). Three decades of multi-dimensional change in global leaf phenology. *Nat. Clim. Change* **5** 364–368.

BYRNE, B., BAKER, D. F., BASU, S., BERTOLACCI, M., BOWMAN, K. W., CARROLL, D., CHATTERJEE, A., CHEVALLIER, F., CIAIS, P. et al. (2023). National CO_2 budgets (2015–2020) inferred from atmospheric CO_2 observations in support of the global stocktake. *Earth Syst. Sci. Data* **15** 963–1004.

CARON, R. J., MCDONALD, J. F. and PONIC, C. M. (1989). A degenerate extreme point strategy for the classification of linear constraints as redundant or necessary. *J. Optim. Theory Appl.* **62** 225–237.

CHEVALLIER, F., FISHER, M., PEYLIN, P., SERRAR, S., BOUSQUET, P., BRÉON, F.-M., CHÉDIN, A. and CIAIS, P. (2005). Inferring CO_2 sources and sinks from satellite observations: Method and application to TOVS data. *J. Geophys. Res., Atmos.* **110**.

CRESSIE, N. and JOHANNESSON, G. (2008). Fixed rank kriging for very large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 209–226.

CROWELL, S., BAKER, D., SCHUH, A., BASU, S., JACOBSON, A. R., CHEVALLIER, F., LIU, J., DENG, F., FENG, L. et al. (2019). The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network. *Atmos. Chem. Phys.* **19** 9797–9831.

FORKEL, M., CARVALHAIS, N., RÖDENBECK, C., KEELING, R., HEIMANN, M., THONICKE, K., ZAEHLE, S. and REICHSTEIN, M. (2016). Enhanced seasonal CO_2 exchange caused by amplified plant productivity in northern ecosystems. *Science* **351** 696–699.

FRIEDLINGSTEIN, P. (2015). Carbon cycle feedbacks and future climate change. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **373**.

GATTI, L. V., BASSO, L. S., MILLER, J. B., GLOOR, M., GATTI DOMINGUES, L., CASSOL, H. L. G., TEJADA, G., ARAGÃO, L. E. O. C., NOBRE, C. et al. (2021). Amazonia as a carbon source linked to deforestation and climate change. *Nature* **595** 388–393.

GELFAND, A. E., KIM, H.-J., SIRMANS, C. F. and BANERJEE, S. (2003). Spatial modeling with spatially varying coefficient processes. *J. Amer. Statist. Assoc.* **98** 387–396.

GONSAMO, A., CHEN, J. M. and OOI, Y. W. (2018). Peak season plant activity shift towards spring is reflected by increasing carbon uptake by extratropical ecosystems. *Glob. Change Biol.* **24** 2117–2128.

GRAVEN, H. D., KEELING, R. F., PIPER, S. C., PATRA, P. K., STEPHENS, B. B., WOFSY, S. C., WELP, L. R., SWEENEY, C., TANS, P. P. et al. (2013). Enhanced seasonal exchange of CO_2 by northern ecosystems since 1960. *Science* **341** 1085–1089.

GURNEY, K. R., LAW, R. M., DENNING, A. S., RAYNER, P. J., BAKER, D., BOUSQUET, P., BRUHWILER, L., CHEN, Y.-H., CIAIS, P. et al. (2002). Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models. *Nature* **415** 626–630.

HAYNES, K. D., BAKER, I. T., DENNING, A. S., STÖCKLI, R., SCHAEFER, K., LOKUPITIYA, E. Y. and HAYNES, J. M. (2019a). Representing grasslands using dynamic prognostic phenology based on biological growth stages: 1. Implementation in the Simple Biosphere Model (SiB4). *J. Adv. Model. Earth Syst.* **11** 4423–4439.

HAYNES, K. D., BAKER, I. T., DENNING, A. S., WOLF, S., WOHLFAHRT, G., KIELY, G., MINAYA, R. C. and HAYNES, J. M. (2019b). Representing grasslands using dynamic prognostic phenology based on biological growth stages: Part 2. Carbon cycling. *J. Adv. Model. Earth Syst.* **11** 4440–4465.

HE, W., JIANG, F., WU, M., JU, W., SCHOLZE, M., CHEN, J. M., BYRNE, B., LIU, J., WANG, H. et al. (2022). China’s terrestrial carbon sink over 2010–2015 constrained by satellite observations of atmospheric CO_2 and land surface variables. *J. Geophys. Res., Biogeosci.* **127** e2021JG006644.

HU, L., MONTZKA, S. A., KAUSHIK, A., ANDREWS, A. E., SWEENEY, C., MILLER, J., BAKER, I. T., DENNING, S., CAMPBELL, E. et al. (2021). COS-derived GPP relationships with temperature and light help explain high-latitude atmospheric CO_2 seasonal cycle amplification. *Proc. Natl. Acad. Sci. USA* **118** e2103423118.

JANJIĆ, T., BORMANN, N., BOCQUET, M., CARTON, J. A., COHN, S. E., DANCE, S. L., LOSA, S. N., NICHOLS, N. K., POTTHAST, R. et al. (2018). On the representation error in data assimilation. *Q. J. R. Meteorol. Soc.* **144** 1257–1278.

KAMINSKI, T., RAYNER, P. J., HEIMANN, M. and ENTING, I. G. (2001). On aggregation errors in atmospheric transport inversions. *J. Geophys. Res., Atmos.* **106** 4703–4715.

KOOIJMANS, L. M. J., CHO, A., MA, J., KAUSHIK, A., HAYNES, K. D., BAKER, I., LUIJKX, I. T., GROENINK, M., PETERS, W. et al. (2021). Evaluation of carbonyl sulfide biosphere exchange in the Simple Biosphere Model (SiB4). *Biogeosciences* **18** 6547–6565.

LANDSCHÜTZER, P., GRUBER, N. and BAKKER, D. C. E. (2016). Decadal variations and trends of the global ocean carbon sink. *Glob. Biogeochem. Cycles* **30** 1396–1417.

LIU, J., BOWMAN, K. W., SCHIMEL, D. S., PARAZOO, N. C., JIANG, Z., LEE, M., BLOOM, A. A., WUNCH, D., FRANKENBERG, C. et al. (2017). Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño. *Science* **358**.

MA, J., KOOIJMANS, L. M. J., CHO, A., MONTZKA, S. A., GLATTHOR, N., WORDEN, J. R., KUAI, L., ATLAS, E. L. and KROL, M. C. (2021). Inverse modelling of carbonyl sulfide: Implementation, evaluation and implications for the global budget. *Atmos. Chem. Phys.* **21** 3507–3529.

MAGNEY, T. S., BARNES, M. L. and YANG, X. (2020). On the covariation of chlorophyll fluorescence and photosynthesis across scales. *Geophys. Res. Lett.* **47** e2020GL091098.

MASSON-DELMOTTE, V., ZHAI, P., PIRANI, A., CONNORS, S. L., PÉAN, C., BERGER, S., CAUD, N., CHEN, Y., GOLDFARB, L. et al. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge Univ. Press, Cambridge, UK.

NORTH, J. S., SCHLIEP, E. M. and WIKLE, C. K. (2021). On the spatial and temporal shift in the archetypal seasonal temperature cycle as driven by annual and semi-annual harmonics. *Environmetrics* **32** e2665.

O'DELL, C. W., ELDERING, A., WENNBERG, P. O., CRISP, D., GUNSON, M. R., FISHER, B., FRANKENBERG, C., KIEL, M., LINDQVIST, H. et al. (2018). Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm. *Atmos. Meas. Tech.* **11** 6539–6576.

PAKMAN, A. and PANINSKI, L. (2014). Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *J. Comput. Graph. Statist.* **23** 518–542.

PALMER, P. I., FENG, L., BAKER, D., CHEVALLIER, F., BÖSCH, H. and SOMKUTI, P. (2019). Net carbon emissions from African biosphere dominate pan-tropical atmospheric CO. *Nat. Commun.* **10** 3344.

PARK, T., CHEN, C., MACIAS-FAURIA, M., TØMMERVIK, H., CHOI, S., WINKLER, A., BHATT, U. S., WALKER, D. A., PIAO, S. et al. (2019). Changes in timing of seasonal peak photosynthetic activity in northern ecosystems. *Glob. Change Biol.* **25** 2382–2395.

PEIRO, H., CROWELL, S., SCHUH, A., BAKER, D. F., O'DELL, C., JACOBSON, A. R., CHEVALLIER, F., LIU, J., ELDERING, A. et al. (2022). Four years of global carbon cycle observed from the Orbiting Carbon Observatory 2 (OCO-2) version 9 and in situ data and comparison to OCO-2 version 7. *Atmos. Chem. Phys.* **22** 1097–1130.

RAYNER, P. J., SCHOLZE, M., KNORR, W., KAMINSKI, T., GIERING, R. and WIDMANN, H. (2005). Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS). *Glob. Biogeochem. Cycles* **19**.

SCHOLZE, M., KAMINSKI, T., KNORR, W., BLESSING, S., VOSSBECK, M., GRANT, J. P. and SCIPAL, K. (2016). Simultaneous assimilation of SMOS soil moisture and atmospheric CO_2 in-situ observations to constrain the global terrestrial carbon cycle. *Remote Sens. Environ.* **180** 334–345.

SCHUH, A. E., BYRNE, B., JACOBSON, A. R., CROWELL, S. M. R., DENG, F., BAKER, D. F., JOHNSON, M. S., PHILIP, S. and WEIR, B. (2022). On the role of atmospheric model transport uncertainty in estimating the Chinese land carbon sink. *Nature* **603** E13–E14.

SCHUH, A. E. and JACOBSON, A. R. (2022). Uncertainty in parameterized convection remains a key obstacle for estimating surface fluxes of carbon dioxide. *Atmos. Chem. Phys. Discuss.* 1–18.

SCHUH, A. E., JACOBSON, A. R., BASU, S., WEIR, B., BAKER, D., BOWMAN, K., CHEVALLIER, F., CROWELL, S., DAVIS, K. J. et al. (2019). Quantifying the impact of atmospheric transport uncertainty on CO_2 surface flux estimates. *Glob. Biogeochem. Cycles* **33** 484–500.

SHEVLIAKOVA, E., STOUFFER, R. J., MALYSHEV, S., KRASTING, J. P., HURTT, G. C. and PACALA, S. W. (2013). Historical warming reduced due to enhanced land carbon uptake. *Proc. Natl. Acad. Sci. USA* **110** 16730–16735.

STELL, A. C., BERTOLACCI, M., ZAMMIT-MANGION, A., RIGBY, M., FRASER, P. J., HARTH, C. M., KRUMMEL, P. B., LAN, X., MANIZZA, M. et al. (2022). Modelling the growth of atmospheric nitrous oxide using a global hierarchical inversion. *Atmos. Chem. Phys.* **22** 12945–12960.

SUN, Y., FRANKENBERG, C., JUNG, M., JOINER, J., GUANTER, L., KÖHLER, P. and MAGNEY, T. (2018). Overview of solar-induced chlorophyll fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, cross-mission comparison, and global monitoring for GPP. *Remote Sens. Environ.* **209** 808–823.

TAGESSON, T., SCHURGERS, G., HORION, S., CIAIS, P., TIAN, F., BRANDT, M., AHLSTRÖM, A., WIGNERON, J.-P., ARDÖ, J. et al. (2020). Recent divergence in the contributions of tropical and boreal forests to the terrestrial carbon sink. *Nat. Ecol. Evol.* **4** 202–209.

TURNER, A. J. and JACOB, D. J. (2015). Balancing aggregation and smoothing errors in inverse models. *Atmos. Chem. Phys.* **15** 7039–7048.

VITOUSEK, P. M., ABER, J. D., HOWARTH, R. W., LIKENS, G. E., MATSON, P. A., SCHINDLER, D. W., SCHLESINGER, W. H. and TILMAN, D. G. (1997). Human alteration of the global nitrogen cycle: Sources and consequences. *Ecol. Appl.* **7** 737–750.

WIKLE, C. K., BERLINER, L. M. and CRESSIE, N. (1998). Hierarchical Bayesian space-time models. *Environ. Ecol. Stat.* **5** 117–154.

WOOD, W. B. (1990). Tropical deforestation: Balancing regional development demands and global environmental concerns. *Glob. Environ. Change* **1** 23–41.

ZAMMIT-MANGION, A., BERTOLACCI, M., FISHER, J., STAVERT, A., RIGBY, M., CAO, Y. and CRESSIE, N. (2022). WOMBAT v1.0: A fully Bayesian global flux-inversion framework. *Geosci. Model Dev.* **15** 45–73.

ZHU, W., TIAN, H., XU, X., PAN, Y., CHEN, G. and LIN, W. (2012). Extension of the growing season due to delayed autumn over mid and high latitudes in North America during 1982–2006. *Glob. Ecol. Biogeogr.* **21** 260–271.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 328–349
https://doi.org/10.1214/23-AOAS1791

ANOPOW FOR REPLICATED NONSTATIONARY TIME SERIES IN EXPERIMENTS

BY ZEDA LI[1,a], YU (RYAN) YUE[1,b] AND SCOTT A. BRUCE[2,c]

[1]*Baruch College, The City University of New York,* [a]*zeda.li@baruch.cuny.edu,* [b]*yu.yue@baruch.cuny.edu*
[2]*Department of Statistics, Texas A&M University,* [c]*sabruce@tamu.edu*

We propose a novel analysis of power (ANOPOW) model for analyzing replicated nonstationary time series commonly encountered in experimental studies. Based on a locally stationary ANOPOW Cramér spectral representation, the proposed model can be used to compare the second-order time-varying frequency patterns among different groups of time series and to estimate group effects as functions of both time and frequency. Formulated in a Bayesian framework, independent two-dimensional second-order random walk (RW2D) priors are assumed on each of the time-varying functional effects for flexible and adaptive smoothing. A piecewise stationary approximation of the nonstationary time series is used to obtain localized estimates of time-varying spectra. Posterior distributions of the time-varying functional group effects are then obtained via integrated nested Laplace approximations (INLA) at a low computational cost. The large-sample distribution of local periodograms can be appropriately utilized to improve estimation accuracy since INLA allows modeling of data with various types of distributions. The usefulness of the proposed model is illustrated through two real-data applications: analyses of seismic signals and pupil diameter time series in children with attention deficit hyperactivity disorder. Simulation studies, Supplementary Material (Li, Yue and Bruce (2024a)), and R code (Li, Yue and Bruce (2024b)) for this article are also available.

REFERENCES

ADAK, S. (1998). Time-dependent spectral analysis of nonstationary time series. *J. Amer. Statist. Assoc.* **93** 1488–1501. MR1666643 https://doi.org/10.2307/2670062

BOLIN, D. and LINDGREN, F. (2015). Excursion and contour uncertainty regions for latent Gaussian models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 85–106. MR3299400 https://doi.org/10.1111/rssb.12055

BOLIN, D. and LINDGREN, F. (2018). Calculating probabilistic excursion sets and related quantities using excursions. *J. Stat. Softw.* **86** 1–20. https://doi.org/10.18637/jss.v086.i05

BRILLINGER, D. R. (2001). *Time Series*: *Data Analysis and Theory*. *Classics in Applied Mathematics* **36**. SIAM, Philadelphia, PA. Reprint of the 1981 edition. MR1853554 https://doi.org/10.1137/1.9780898719246

BRUCE, S. A., HALL, M. H., BUYSSE, D. J. and KRAFTY, R. T. (2018). Conditional adaptive Bayesian spectral analysis of nonstationary biomedical time series. *Biometrics* **74** 260–269. MR3777946 https://doi.org/10.1111/biom.12719

CHAU, J. and VON SACHS, R. (2016). Functional mixed effects wavelet estimation for spectra of replicated time series. *Electron. J. Stat.* **10** 2461–2510. MR3545466 https://doi.org/10.1214/16-EJS1181

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916 https://doi.org/10.1214/aos/1034276620

DIGGLE, P. J. and AL WASEL, I. (1997). Spectral analysis of replicated biomedical time series. *J. R. Stat. Soc., Ser. C* **46** 31–71. With discussion and a reply by the authors. MR1452286 https://doi.org/10.1111/1467-9876.00047

DUCHOWSKI, A. T., KREJTZ, K., GEHRER, N. A., BAFNA, T. and BÆKGAARD, P. (2020). The low/high index of pupillary activity. In *Proceedings of the* 2020 *CHI Conference on Human Factors in Computing Systems* 1–12.

Key words and phrases. Integrated nested Laplace approximations, locally stationary ANOPOW Cramér spectral representation, pupil diameter time series, replicated nonstationary time series, spectral analysis.

FIECAS, M. and OMBAO, H. (2016). Modeling the evolution of dynamic brain processes during an associative learning experiment. *J. Amer. Statist. Assoc.* **111** 1440–1453. MR3601700 https://doi.org/10.1080/01621459.2016.1165683

FREYERMUTH, J.-M., OMBAO, H. and VON SACHS, R. (2010). Tree-structured wavelet estimation in a mixed effects model for spectra of replicated time series. *J. Amer. Statist. Assoc.* **105** 634–646. MR2724848 https://doi.org/10.1198/jasa.2010.tm09132

FRYZLEWICZ, P. and OMBAO, H. (2009). Consistent classification of nonstationary time series using stochastic wavelet representations. *J. Amer. Statist. Assoc.* **104** 299–312. MR2504379 https://doi.org/10.1198/jasa.2009.0110

GABRY, J., SIMPSON, D., VEHTARI, A., BETANCOURT, M. and GELMAN, A. (2019). Visualization in Bayesian workflow. *J. Roy. Statist. Soc. Ser. A* **182** 389–402. MR3902665 https://doi.org/10.1111/rssa.12378

GUO, W. and DAI, M. (2006). Multivariate time-dependent spectral analysis using Cholesky decomposition. *Statist. Sinica* **16** 825–845. MR2281304

GUO, W., DAI, M., OMBAO, H. C. and VON SACHS, R. (2003). Smoothing spline ANOVA for time-dependent spectral analysis. *J. Amer. Statist. Assoc.* **98** 643–652. MR2011677 https://doi.org/10.1198/016214503000000549

HAMED, A. M., KAUER, A. J. and STEVENS, H. E. (2015). Why the diagnosis of attention deficit hyperactivity disorder matters. *Front. Psychiatry* **6** 168. https://doi.org/10.3389/fpsyt.2015.00168

HELD, L., SCHRÖDLE, B. and RUE, H. (2010). Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA. In *Statistical Modelling and Regression Structures* 91–110. Physica-Verlag, Heidelberg. MR2664630 https://doi.org/10.1007/978-3-7908-2413-1_6

IANNACCONE, R. and COLES, S. (2001). Semiparametric models and inference for biomedical time series with extra-variation. *Biostatistics* **2** 261–276.

KLEY, T., VOLGUSHEV, S., DETTE, H. and HALLIN, M. (2016). Quantile spectral processes: Asymptotic analysis and inference. *Bernoulli* **22** 1770–1807. MR3474833 https://doi.org/10.3150/15-BEJ711

KRAFTY, R. T. and COLLINGE, W. O. (2013). Penalized multivariate Whittle likelihood for power spectrum estimation. *Biometrika* **100** 447–458. MR3068445 https://doi.org/10.1093/biomet/ass088

KRAFTY, R. T., HALL, M. and GUO, W. (2011). Functional mixed effects spectral analysis. *Biometrika* **98** 583–598. MR2836408 https://doi.org/10.1093/biomet/asr032

LI, Z. (2023). Robust conditional spectral analysis of replicated time series. *Stat. Interface* **16** 81–96. MR4465931

LI, Z., BRUCE, S. A., WUTZKE, C. J. and LONG, Y. (2021). Conditional adaptive Bayesian spectral analysis of replicated multivariate time series. *Stat. Med.* **40** 1989–2005. MR4229834 https://doi.org/10.1002/sim.8884

LI, Z. and KRAFTY, R. T. (2019). Adaptive Bayesian time-frequency analysis of multivariate time series. *J. Amer. Statist. Assoc.* **114** 453–465. MR3941268 https://doi.org/10.1080/01621459.2017.1415908

LI, Z., YUE, Y. (R.) and BRUCE, S. A. (2024a). Supplementary materials to "ANOPOW for replicated nonstationary time series in experiments." https://doi.org/10.1214/23-AOAS1791SUPPA.

LI, Z., YUE, Y. (R.) and BRUCE, S. A. (2024b). R code for "ANOPOW for replicated nonstationary time series in experiments." https://doi.org/10.1214/23-AOAS1791SUPPB.

MARTINEZ, J. G., BOHN, K. M., CARROLL, R. J. and MORRIS, J. S. (2013). A study of Mexican free-tailed bat chirp syllables: Bayesian functional mixed models for nonstationary acoustic time series. *J. Amer. Statist. Assoc.* **108** 514–526. MR3174638 https://doi.org/10.1080/01621459.2013.793118

MARTINS, T. G., SIMPSON, D., LINDGREN, F. and RUE, H. (2013). Bayesian computing with INLA: New features. *Comput. Statist. Data Anal.* **67** 68–83. MR3079584 https://doi.org/10.1016/j.csda.2013.04.014

MATHÔT, S. (2018). Pupillometry: Psychology, physiology, and function. *J. Cogn.* **1**.

MORRIS, J. S. and CARROLL, R. J. (2006). Wavelet-based functional mixed models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 179–199. MR2188981 https://doi.org/10.1111/j.1467-9868.2006.00539.x

NAKAYAMA, M. and SHIMIZU, Y. (2004). Frequency analysis of task evoked pupillary response and eye-movement. In *Proceedings of the* 2004 *Symposium on Eye Tracking Research and Applications*. *ETRA'*04 71–76. Assoc. Comput. Mach., New York, NY, USA. https://doi.org/10.1145/968363.968381

QIN, L., GUO, W. and LITT, B. (2009). A time-frequency functional model for locally stationary time series data. *J. Comput. Graph. Statist.* **18** 675–693. MR2572632 https://doi.org/10.1198/jcgs.2009.06109

R CORE TEAM (2023). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

REIMER, J., FROUDARAKIS, E., CADWELL, C. R., YATSENKO, D., DENFIELD, G. H. and TOLIAS, A. S. (2014). Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron* **84** 355–362.

ROJAS-LÍBANO, D., WAINSTEIN, G., CARRASCO, X., ABOITIZ, F. and CROSSLEY, N. (2019). A pupil size, eye-tracking and neuropsychological dataset from ADHD children during a cognitive task. *Sci. Data* **25**. https://doi.org/10.1038/s41597-019-0037-2

ROSEN, O., WOOD, S. and STOFFER, D. S. (2012). AdaptSPEC: Adaptive spectral estimation for nonstationary time series. *J. Amer. Statist. Assoc.* **107** 1575–1589. MR3036417 https://doi.org/10.1080/01621459.2012.716340

RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications. Monographs on Statistics and Applied Probability* **104**. CRC Press, Boca Raton, FL. MR2130347 https://doi.org/10.1201/9780203492024

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

SHUMWAY, R. H. and STOFFER, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*, 4th ed. *Springer Texts in Statistics*. Springer, Cham. MR3642322 https://doi.org/10.1007/978-3-319-52452-8

SØRBYE, S. H., HINDBERG, K., OLSEN, L. R. and RUE, H. (2009). Bayesian multiscale feature detection of log-spectral densities. *Comput. Statist. Data Anal.* **53** 3746–3754. MR2749919 https://doi.org/10.1016/j.csda.2009.03.020

STOFFER, D. and POISON, N. (2023). Astsa: Applied statistical time series analysis. R package version 1.15.

STOFFER, D. S., HAN, S., QIN, L. and GUO, W. (2010). Smoothing spline ANOPOW. *J. Statist. Plann. Inference* **140** 3789–3796. MR2674166 https://doi.org/10.1016/j.jspi.2010.04.043

WAHBA, G. (1980). Automatic smoothing of the log-periodogram. *J. Amer. Statist. Assoc.* **75** 122–132.

WAHBA, G. (1990). *Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442 https://doi.org/10.1137/1.9781611970128

WAINSTEIN, G., ROJAS-LÍBANO, D., CROSSLEY, N. A., CARRASCO, X., ABOITIZ, F. and OSSANDÓN, T. (2017). Pupil size tracks attentional performance in attention-deficit/hyperactivity disorder. *Sci. Rep.* **8228**. https://doi.org/10.1038/s41597-019-0037-2

WANG, X., YUE, Y. R. and FARAWAY, J. J. (2018). *Bayesian Regression Modeling with INLA. Chapman & Hall/CRC Computer Science and Data Analysis Series*. CRC Press, Boca Raton, FL. MR3752633

WANG, Y., LI, Z. and BRUCE, S. A. (2022). Adaptive Bayesian sum of trees model for covariate-dependent spectral analysis. *Biometrics*. https://doi.org/10.1111/biom.13763

YUE, Y., BOLIN, D., RUE, H. and WANG, X.-F. (2019). Bayesian generalized two-way ANOVA modeling for functional data using INLA. *Statist. Sinica* **29** 741–767. MR3931386

YUE, Y. and SPECKMAN, P. L. (2010). Nonstationary spatial Gaussian Markov random fields. *J. Comput. Graph. Statist.* **19** 96–116. With supplementary data available online. MR2654402 https://doi.org/10.1198/jcgs.2009.08124

ZHANG, S. (2020). Nonparametric Bayesian inference for the spectral density based on irregularly spaced data. *Comput. Statist. Data Anal.* **151** 107019, 14. MR4107729 https://doi.org/10.1016/j.csda.2020.107019

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 350–374
https://doi.org/10.1214/23-AOAS1792

A BAYESIAN MACHINE LEARNING APPROACH FOR ESTIMATING HETEROGENEOUS SURVIVOR CAUSAL EFFECTS: APPLICATIONS TO A CRITICAL CARE TRIAL

BY XINYUAN CHEN[1,a], MICHAEL O. HARHAY[2,b], GUANGYU TONG[3,c] AND FAN LI[3,d]

[1]*Department of Mathematics and Statistics, Mississippi State University,* [a]*xchen@math.msstate.edu*
[2]*Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania,* [b]*mharhay@pennmedicine.upenn.edu*
[3]*Department of Biostatistics, Yale School of Public Health,* [c]*guangyu.tong@yale.edu,* [d]*fan.f.li@yale.edu*

Assessing heterogeneity in the effects of treatments has become increasingly popular in the field of causal inference and carries important implications for clinical decision-making. While extensive literature exists for studying treatment effect heterogeneity when outcomes are fully observed, there has been limited development in tools for estimating heterogeneous causal effects when patient-centered outcomes are truncated by a terminal event, such as death. Due to mortality occurring during study follow-up, the outcomes of interest are unobservable, undefined, or not fully observed for many participants in which case principal stratification is an appealing framework to draw valid causal conclusions. Motivated by the Acute Respiratory Distress Syndrome Network (ARDSNetwork) ARDS respiratory management (ARMA) trial, we developed a flexible Bayesian machine learning approach to estimate the average causal effect and heterogeneous causal effects among the always-survivors stratum when clinical outcomes are subject to truncation. We adopted Bayesian additive regression trees (BART) to flexibly specify separate mean models for the potential outcomes and latent stratum membership. In the analysis of the ARMA trial, we found that the low tidal volume treatment had an overall benefit for participants sustaining acute lung injuries on the outcome of time to returning home but substantial heterogeneity in treatment effects among the always-survivors, driven most strongly by biologic sex and the alveolar-arterial oxygen gradient at baseline (a physiologic measure of lung function and degree of hypoxemia). These findings illustrate how the proposed methodology could guide the prognostic enrichment of future trials in the field.

REFERENCES

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

AUSTIN, P. C. and STUART, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat. Med.* **34** 3661–3679. MR3422140 https://doi.org/10.1002/sim.6607

BARGAGLI-STOFFI, F. J., DE WITTE, K. and GNECCO, G. (2022). Heterogeneous causal effects with imperfect compliance: A Bayesian machine learning approach. *Ann. Appl. Stat.* **16** 1986–2009. MR4455908 https://doi.org/10.1214/21-aoas1579

BIA, M., MATTEI, A. and MERCATANTI, A. (2022). Assessing causal effects in a longitudinal observational study with “truncated” outcomes due to unemployment and nonignorable missing data. *J. Bus. Econom. Statist.* **40** 718–729. MR4410893 https://doi.org/10.1080/07350015.2020.1862672

BROWER, R. G., MATTHAY, M. A., MORRIS, A., SCHOENFELD, D., THOMPSON, B. T., WHEELER, A. et al. (2000). Acute respiratory distress syndrome network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N. Engl. J. Med.* **342** 1301–1308.

Key words and phrases. Acute lung injury, Bayesian additive regression trees, causal inference, heterogeneity of treatment effects, principal stratification, truncation by death.

CHEN, X., HARHAY, M. O., TONG, G. and LI, F. (2024). Supplement to “A Bayesian machine learning approach for estimating heterogeneous survivor causal effects: Applications to a critical care trial.” https://doi.org/10.1214/23-AOAS1792SUPPA, https://doi.org/10.1214/23-AOAS1792SUPPB

CHIBA, Y. and VANDERWEELE, T. J. (2011). A simple method for principal strata effects when the outcome has been truncated due to death. *Amer. J. Epidemiol.* **173** 745–751. https://doi.org/10.1093/aje/kwq418

CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4** 266–298. MR2758172 https://doi.org/10.1214/09-AOAS285

COMMENT, L., MEALLI, F., HANEUSE, S. and ZIGLER, C. (2019). Survivor average causal effects for continuous time: A principal stratification approach to causal inference with semicompeting risks. arXiv preprint arXiv:1902.09304, 1–28.

DEL SORBO, L., GOLIGHER, E. C., MCAULEY, D. F., RUBENFELD, G. D., BROCHARD, L. J., GATTINONI, L., SLUTSKY, A. S. and FAN, E. (2017). Mechanical ventilation in adults with acute respiratory distress syndrome. Summary of the experimental evidence for the clinical practice guideline. *Ann. Amer. Thorac. Soc.* **14** S261–S270. PMID: 28985479. https://doi.org/10.1513/AnnalsATS.201704-345OT

DENG, Y., GUO, Y., CHANG, Y. and ZHOU, X.-H. (2021). Identification and estimation of the heterogeneous survivor average causal effect in observational studies. arXiv preprint arXiv:2109.13623, 1–23.

DING, P., GENG, Z., YAN, W. and ZHOU, X.-H. (2011). Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *J. Amer. Statist. Assoc.* **106** 1578–1591. MR2896858 https://doi.org/10.1198/jasa.2011.tm10265

DING, P. and LI, F. (2018). Causal inference: A missing data perspective. *Statist. Sci.* **33** 214–237. MR3797711 https://doi.org/10.1214/18-STS645

DING, P. and LU, J. (2017). Principal stratification analysis using principal scores. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 757–777. MR3641406 https://doi.org/10.1111/rssb.12191

DORIE, V., HILL, J., SHALIT, U., SCOTT, M. and CERVONE, D. (2019). Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statist. Sci.* **34** 43–68. MR3938963 https://doi.org/10.1214/18-STS667

EGLESTON, B. L., SCHARFSTEIN, D. O., FREEMAN, E. E. and WEST, S. K. (2006). Causal inference for non-mortality outcomes in the presence of death. *Biostatistics* **8** 526–545. https://doi.org/10.1093/biostatistics/kxl027

FAN, E., DEL SORBO, L., GOLIGHER, E. C., HODGSON, C. L., MUNSHI, L., WALKEY, A. J., ADHIKARI, N. K. J., AMATO, M. B. P., BRANSON, R. et al. (2017). An official American thoracic society/European society of intensive care medicine/society of critical care medicine clinical practice guideline: Mechanical ventilation in adult patients with acute respiratory distress syndrome. *Am. J. Respir. Crit. Care Med.* **195** 1253–1263. PMID: 28459336. https://doi.org/10.1164/rccm.201703-0548ST

FRANGAKIS, C. E. and RUBIN, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58** 21–29. MR1891039 https://doi.org/10.1111/j.0006-341X.2002.00021.x

FRUMENTO, P., MEALLI, F., PACINI, B. and RUBIN, D. B. (2012). Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data. *J. Amer. Statist. Assoc.* **107** 450–466. MR2980057 https://doi.org/10.1080/01621459.2011.643719

GOLIGHER, E. C., COSTA, E. L. V., YARNELL, C. J., BROCHARD, L. J., STEWART, T. E., TOMLINSON, G., BROWER, R. G., SLUTSKY, A. S. and AMATO, M. P. B. (2021). Effect of lowering Vt on mortality in acute respiratory distress syndrome varies with respiratory system elastance. *Am. J. Respir. Crit. Care Med.* **203** 1378–1385. https://doi.org/10.1164/rccm.202009-3536OC

HAHN, P. R., MURRAY, J. S. and CARVALHO, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Anal.* **15** 965–1056. Includes comments and discussions by 25 discussants and a rejoinder by the authors. MR4154846 https://doi.org/10.1214/19-BA1195

HAHN, P. R., MURRAY, J. S. and MANOLOPOULOU, I. (2016). A Bayesian partial identification approach to inferring the prevalence of accounting misconduct. *J. Amer. Statist. Assoc.* **111** 14–26. MR3494635 https://doi.org/10.1080/01621459.2015.1084307

HARHAY, M. O., RATCLIFFE, S. J., SMALL, D. S., SUTTNER, L. H., CROWTHER, M. J. and HALPERN, S. D. (2019). Measuring and analyzing length of stay in critical care trials. *Med. Care* **57** e53–e59. PMID: 30664613. https://doi.org/10.1097/MLR.0000000000001059

HARHAY, M. O., WAGNER, J., RATCLIFFE, S. J., BRONHEIM, R. S., GOPAL, A., GREEN, S., COONEY, E., MIKKELSEN, M. E., KERLIN, M. P. et al. (2014). Outcomes and statistical power in adult critical care randomized trials. *Am. J. Respir. Crit. Care Med.* **189** 1469–1478.

HAYDEN, D., PAULER, D. K. and SCHOENFELD, D. (2005). An estimator for treatment comparisons among survivors in randomized trials. *Biometrics* **61** 305–310. MR2135873 https://doi.org/10.1111/j.0006-341X.2005.030227.x

HELMHOLZ JR., H. F. (1979). The abbreviated alveolar air equation. *Chest* **75** 748. https://doi.org/10.1378/chest.75.6.748

HENDERSON, N. C., LOUIS, T. A., ROSNER, G. L. and VARADHAN, R. (2020). Individualized treatment effects with censored data via fully nonparametric Bayesian accelerated failure time models. *Biostatistics* **21** 50–68. MR4043845 https://doi.org/10.1093/biostatistics/kxy028

HILL, J. L. (2011). Bayesian nonparametric modeling for causal inference. *J. Comput. Graph. Statist.* **20** 217–240. Supplementary material available online. MR2816546 https://doi.org/10.1198/jcgs.2010.08162

HIRANO, K., IMBENS, G. W., RUBIN, D. B. and ZHOU, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* **1** 69–88.

HU, L., JI, J. and LI, F. (2021). Estimating heterogeneous survival treatment effect in observational data using machine learning. *Stat. Med.* **40** 4691–4713. MR4315446 https://doi.org/10.1002/sim.9090

IMAI, K. (2008). Sharp bounds on the causal effects in randomized experiments with “truncation-by-death”. *Statist. Probab. Lett.* **78** 144–149. MR2382067 https://doi.org/10.1016/j.spl.2007.05.015

KADANE, J. B. (1975). The role of identification in Bayesian theory. In *Studies in Bayesian Econometrics and Statistics* (*in Honor of Leonard J. Savage*). *Contrib. Econom. Anal.* **86** 175–191. North-Holland, Amsterdam–Oxford. MR0483124

KIM, C., DANIELS, M. J., HOGAN, J. W., CHOIRAT, C. and ZIGLER, C. M. (2019). Bayesian methods for multiple mediators: Relating principal stratification and causal mediation in the analysis of power plant emission controls. *Ann. Appl. Stat.* **13** 1927–1956. MR4019162 https://doi.org/10.1214/19-AOAS1260

KIM, C., DANIELS, M. J., MARCUS, B. H. and ROY, J. A. (2017). A framework for Bayesian nonparametric inference for causal effects of mediation. *Biometrics* **73** 401–409. MR3665957 https://doi.org/10.1111/biom.12575

LI, F. and LI, F. (2019). Propensity score weighting for causal inference with multiple treatments. *Ann. Appl. Stat.* **13** 2389–2415. MR4037435 https://doi.org/10.1214/19-aoas1282

LONG, D. M. and HUDGENS, M. G. (2013). Sharpening bounds on principal effects with covariates. *Biometrics* **69** 812–819. MR3146777 https://doi.org/10.1111/biom.12103

MATTEI, A., LI, F. and MEALLI, F. (2013). Exploiting multiple outcomes in Bayesian principal stratification analysis with application to the evaluation of a job training program. *Ann. Appl. Stat.* **7** 2336–2360. MR3161725 https://doi.org/10.1214/13-AOAS674

MATTEI, A. and MEALLI, F. (2007). Application of the principal stratification approach to the Faenza randomized experiment on breast self-examination. *Biometrics* **63** 437–446. MR2370802 https://doi.org/10.1111/j.1541-0420.2006.00684.x

MATTHAY, M. A., MCAULEY, D. F. and WARE, L. B. (2017). Clinical trials in acute respiratory distress syndrome: Challenges and opportunities. *Lancet Respir. Med.* **5** 524–534. https://doi.org/10.1016/S2213-2600(17)30188-1

MCCAFFREY, D. F., GRIFFIN, B. A., ALMIRALL, D., SLAUGHTER, M. E., RAMCHAND, R. and BURGETTE, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Stat. Med.* **32** 3388–3414. MR3074364 https://doi.org/10.1002/sim.5753

MCNICHOLAS, B. A., MADOTTO, F., PHAM, T., REZOAGLI, E., MASTERSON, C. H., HORIE, S., BELLANI, G., BROCHARD, L. and LAFFEY, J. G. (2019). Demographics, management and outcome of women and men with acute respiratory distress syndrome in the LUNG SAFE prospective cohort study. *Eur. Respir. J.* **54**. PMID: 31346004. https://doi.org/10.1183/13993003.00609-2019

NEVO, D. and GORFINE, M. (2022). Causal inference for semi-competing risks data. *Biostatistics* **23** 1115–1132. MR4496371 https://doi.org/10.1093/biostatistics/kxab049

PAPAKOSTAS, D., HAHN, P. R., MURRAY, J., ZHOU, F. and GERAKOS, J. (2023). Do forecasts of bankruptcy cause bankruptcy? A machine learning sensitivity analysis. *Ann. Appl. Stat.* **17** 711–739. MR4539050 https://doi.org/10.1214/22-aoas1648

ROY, J., LUM, K. J. and DANIELS, M. J. (2017). A Bayesian nonparametric approach to marginal structural models for point treatments and a continuous or survival outcome. *Biostatistics* **18** 32–47. MR3612272 https://doi.org/10.1093/biostatistics/kxw029

SHEN, Y., CAI, G., GONG, S., DONG, L., YAN, J. and CAI, W. (2019). Interaction between low tidal volume ventilation strategy and severity of acute respiratory distress syndrome: A retrospective cohort study. *Crit. Care* **23** 254. https://doi.org/10.1186/s13054-019-2530-6

TAN, Y. V. and ROY, J. (2019). Bayesian additive regression trees and the general BART model. *Stat. Med.* **38** 5048–5069. MR4022845 https://doi.org/10.1002/sim.8347

TONELLI, A. R., ZEIN, J., ADAMS, J. and IOANNIDIS, J. P. A. (2014). Effects of interventions on survival in acute respiratory distress syndrome: An umbrella review of 159 published randomized trials and 29 meta-analyses. *Intens. Care Med.* **40** 769–787. https://doi.org/10.1007/s00134-014-3272-1

WANG, L., ZHOU, X.-H. and RICHARDSON, T. S. (2017). Identification and estimation of causal effects with outcomes truncated by death. *Biometrika* **104** 597–612. MR3694585 https://doi.org/10.1093/biomet/asx034

WENDLING, T., JUNG, K., CALLAHAN, A., SCHULER, A., SHAH, N. H. and GALLEGO, B. (2018). Comparing methods for estimation of heterogeneous treatment effects using observational data from health care databases. *Stat. Med.* **37** 3309–3324. MR3856345 https://doi.org/10.1002/sim.7820

WOODY, S., CARVALHO, C. M. and MURRAY, J. S. (2021). Model interpretation through lower-dimensional posterior summarization. *J. Comput. Graph. Statist.* **30** 144–161. MR4235972 https://doi.org/10.1080/10618600.2020.1796684

XU, Y., MÜLLER, P., WAHED, A. S. and THALL, P. F. (2016). Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. *J. Amer. Statist. Assoc.* **111** 921–950. MR3561917 https://doi.org/10.1080/01621459.2015.1086353

XU, Y., SCHARFSTEIN, D., MÜLLER, P. and DANIELS, M. (2022). A Bayesian nonparametric approach for evaluating the causal effect of treatment in randomized trials with semi-competing risks. *Biostatistics* **23** 34–49. MR4366034 https://doi.org/10.1093/biostatistics/kxaa008

YANG, F. and SMALL, D. S. (2016). Using post-outcome measurement information in censoring-by-death problems. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 299–318. MR3453657 https://doi.org/10.1111/rssb.12113

ZHANG, J. L. and RUBIN, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *J. Educ. Behav. Stat.* **28** 353–368. https://doi.org/10.3102/10769986028004353

ZHANG, J. L., RUBIN, D. B. and MEALLI, F. (2009). Likelihood-based analysis of causal effects of job-training programs using principal stratification. *J. Amer. Statist. Assoc.* **104** 166–176. MR2663040 https://doi.org/10.1198/jasa.2009.0012

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 375–403
https://doi.org/10.1214/23-AOAS1793

BAYESIAN PROJECTIONS OF TOTAL FERTILITY RATE CONDITIONAL ON THE UNITED NATIONS SUSTAINABLE DEVELOPMENT GOALS

BY DAPHNE H. LIU[1,a] AND ADRIAN E. RAFTERY[2,b]

[1]*Department of Statistics, University of Washington,* [a]*dhliu@uw.edu*
[2]*Departments of Statistics and Sociology, University of Washington,* [b]*raftery@uw.edu*

Women's educational attainment and contraceptive prevalence are two mechanisms identified as having an accelerating effect on fertility decline and that can be directly impacted by policy. Quantifying the potential accelerating effect of education and family planning policies on fertility decline in a probabilistic way is of interest to policymakers, particularly in high-fertility countries. We propose a conditional Bayesian hierarchical model for projecting fertility, given education and family planning policy interventions. To illustrate the effect policy changes could have on future fertility, we create probabilistic projections of fertility that condition on scenarios such as achieving the sustainable development goals (SDGs) for universal secondary education and universal access to family planning by 2030.

REFERENCES

ABEL, G. J., BARAKAT, B., KC, S. and LUTZ, W. (2016). Meeting the sustainable development goals leads to lower world population growth. *Proc. Natl. Acad. Sci. USA* **113** 14294–14299. https://doi.org/10.1073/pnas.1611386113

ADSERÁ, A. (2017a). Education and fertility in the context of rising inequality. *Vienna Yearb. Popul. Res.* **15** 63–92. https://doi.org/10.1553/populationyearbook2017s063

ADSERÁ, A. (2017b). The future fertility of highly educated women: The role of educational composition shifts and labor market barriers. *Vienna Yearb. Popul. Res.* **15** 19–25. https://doi.org/10.1553/populationyearbook2017s019

ALKEMA, L. (2020). The Global Burden of Disease fertility forecasts: Summary of the approach used and associated statistical concerns. OSF preprints. https://doi.org/10.31219/osf.io/3m6va

ALKEMA, L., RAFTERY, A. E., GERLAND, P., CLARK, S. J., PELLETIER, F., BUETTNER, T. and HEILIG, G. K. (2011). Probabilistic projections of the total fertility rate for all countries. *Demography* **48** 815–839. https://doi.org/10.1007/s13524-011-0040-5

AXINN, W. G. and BARBER, J. S. (2001). Mass education and fertility transition. *Am. Sociol. Rev.* **66** 481–505. https://doi.org/10.2307/3088919

AZOSE, J. J. and RAFTERY, A. E. (2015). Bayesian probabilistic projection of international migration rates. *Demography* **52** 1627–1650. https://doi.org/10.1007/s13524-015-0415-0

BONGAARTS, J. (1978). A framework for analyzing the proximate determinants of fertility. *Popul. Dev. Rev.* **4** 105–132. https://doi.org/10.2307/1972149

BONGAARTS, J. (1987). The proximate determinants of exceptionally high fertility. *Popul. Dev. Rev.* **13** 133–139. https://doi.org/10.2307/1972125

BONGAARTS, J. (2003). Completing the fertility transition in the developing world: The role of educational differences and fertility preferences. *Popul. Stud.* **57** 321–335. https://doi.org/10.1080/0032472032000137835

BONGAARTS, J. (2010). The causes of educational differences in fertility in sub-Saharan Africa. Poverty, gender, and youth. Working Paper no. 20. New York: Population Council. https://doi.org/10.31899/pgy3.1023

BONGAARTS, J. (2013). Demographic trends and implications for development. In *IUSSP* 2013 *Meeting*, *Busan*. Available at https://iussp.org/en/busan-conference-report.

BONGAARTS, J. (2016). Development: Slow down population growth. *Nature* **530** 409–412. https://doi.org/10.1038/530409a

BONGAARTS, J. (2017). The effect of contraception on fertility: Is sub-Saharan Africa different? *Demogr. Res.* **37** 129–146. https://doi.org/10.4054/DemRes.2017.37.6

BONGAARTS, J. and CASTERLINE, J. (2013). Fertility transition: Is sub-Saharan Africa different? *Popul. Dev. Rev.* **38** 153–168. https://doi.org/10.1111/j.1728-4457.2013.00557.x

---

Key words and phrases. Bayesian hierarchical model, fertility, sustainable development goals.

BONGAARTS, J., FRANK, O. and LESTHAEGHE, R. (1984). The proximate determinants of fertility in sub-Saharan Africa. *Popul. Dev. Rev.* **10** 511–537. https://doi.org/10.2307/1973518

BONGAARTS, J., MENSCH, B. S. and BLANC, A. K. (2017). Trends in the age at reproductive transitions in the developing world: The role of education. *Popul. Stud.* **71** 139–154. https://doi.org/10.1080/00324728.2017.1291986

CAHILL, N., WEINBERGER, M. and ALKEMA, L. (2020). What increase in modern contraceptive use is needed in FP2020 countries to reach 75% demand satisfied by 2030? An assessment using the accelerated transition method and family planning estimation model. *Gates Open Res.* **4**. https://doi.org/10.12688/gatesopenres.13125.1

CALDWELL, J. C. (1982). *Theory of Fertility Decline*. Academic Press, New York.

CALDWELL, J. C., REDDY, P. H. and CALDWELL, P. (1985). Educational transition in rural south India. *Popul. Dev. Rev.* **11** 29–51. https://doi.org/10.2307/1973377

DE SILVA, T. and TENREYRO, S. (2017). Population control policies and fertility convergence. *J. Econ. Perspect.* **31** 205–228. https://doi.org/10.1257/jep.31.4.205

EASTERLIN, R. A. and CRIMMINS, E. M. (1985). *The Fertility Revolution*: *A Supply-Demand Analysis*. Univ. Chicago Press, Chicago, IL.

FABIC, M. S., CHOI, Y., BONGAARTS, J., DARROCH, J. E., ROSS, J. A., STOVER, J., TSUI, A. O., UPADHYAY, J. and STARBIRD, E. (2015). Meeting demand for family planning within a generation: The post-2015 agenda. *Lancet* **385** 1928–1931. https://doi.org/10.1016/S0140-6736(14)61055-2

FOSDICK, B. K. and RAFTERY, A. E. (2014). Regional probabilistic fertility forecasting by modeling between-country correlations. *Demogr. Res.* **30** 1011–1034. https://doi.org/10.4054/DemRes.2014.30.35

FRIEDMAN, J., YORK, H., GRAETZ, N., WOYCZYNSKI, L., WHISNANT, J., HAY, S. I. and GAKIDOU, E. (2020). Measuring and forecasting progress towards the education-related SDG targets. *Nature* **580** 636–639. https://doi.org/10.1038/s41586-020-2198-8

GIETEL-BASTEN, S. and SOBOTKA, T. (2020). Uncertain population futures: Critical reflections on the IHME scenarios of future fertility, mortality, migration and population trends from 2017 to 2100. SocArXiv. https://doi.org/10.31235/osf.io/5syef

GIETEL-BASTEN, S. and SOBOTKA, T. (2021). Trends in population health and demography. *Lancet* **398** 580–581. https://doi.org/10.1016/S0140-6736(21)01051-5

GODWIN, J. and RAFTERY, A. E. (2017). Bayesian projection of life expectancy accounting for the HIV/AIDS epidemic. *Demogr. Res.* **37** 1549–1610. https://doi.org/10.4054/DemRes.2017.37.48

GRANT, M. J. (2015). The demographic promise of expanded female education: Trends in the age at first birth in Malawi. *Popul. Dev. Rev.* **41** 409–438. https://doi.org/10.1111/j.1728-4457.2015.00066.x

HERZER, D., STRULIK, H. and VOLLMER, S. (2012). The long-run determinants of fertility: One century of demographic change 1900–1999. *J. Econ. Growth* **17** 357–385. https://doi.org/10.1007/s10887-012-9085-6

HIRSCHMAN, C. (1994). Why fertility changes. *Annu. Rev. Sociol.* **20** 203–233. https://doi.org/10.1146/annurev.so.20.080194.001223

KANEDA, T., FALK, M. and PATIERNO, K. (2021). Understanding and comparing population projections in sub-Saharan Africa. Available at https://www.prb.org/resources/understanding-and-comparing-population-projections-in-sub-saharan-africa/. Accessed: 2021-07-23.

KANTOROVÁ, V., WHELDON, M. C., UEFFING, P. and DASGUPTA, A. N. Z. (2020). Estimating progress towards meeting women's contraceptive needs in 185 countries: A Bayesian hierarchical modelling study. *PLoS Med.* **17** e1003026. https://doi.org/10.1371/journal.pmed.1003026

LEE, R. and MASON, A. (2006). What is the demographic dividend? *Financ. Dev.* **43** 16–24.

LIU, D. H. and RAFTERY, A. E. (2020). How do education and family planning accelerate fertility decline? *Popul. Dev. Rev.* **46** 409–441. https://doi.org/10.1111/padr.12347

LIU, D. H. and RAFTERY, A. E. (2024). Supplement to "Bayesian projections of total fertility rate conditional on the United Nations sustainable development goals." https://doi.org/10.1214/23-AOAS1793SUPP

LIU, P., ŠEVČÍKOVÁ, H. and RAFTERY, A. E. (2023). Probabilistic estimation and projection of the annual total fertility rate accounting for past uncertainty: A major update of the bayesTFR R package. *J. Stat. Softw.* **106** 1–36. https://doi.org/10.18637/jss.v106.i08

LUTZ, W. (2023). Population decline will likely become a global trend and benefit long-term human wellbeing. *Vienna Yearb. Popul. Res.* **21**. https://doi.org/10.1553/p-3cp7-4e6b

LUTZ, W., BUTZ, W. P. and KC, S., eds. (2014). *World Population and Human Capital in the Twenty-First Century* Oxford Univ. Press, Oxford.

LUTZ, W., GOUJON, A., KC, S., STONAWSKI, M. and STILIANAKIS, N., eds. (2018). *Demographic and Human Capital Scenarios for the* 21*st Century*: 2018 *Assessment for* 201 *Countries* Publications Office of the European Union, Luxembourg.

MAATHUIS, M. H. and COLOMBO, D. (2015). A generalized back-door criterion. *Ann. Statist.* **43** 1060–1088. MR3346697 https://doi.org/10.1214/14-AOS1295

MADDISON PROJECT DATABASE, VERSION 2018, BOLT, J., INKLAAR, R., DE JONG, H. and VAN ZANDEN, J. L. (2018). Rebasing 'Maddison': New income comparisons and the shape of long-run economic development. Maddison Project Working paper 10.

MARTÍN, T. C. (1995). Women's education and fertility: Results from 26 Demographic and Health Surveys. *Stud. Fam. Plan.* **26** 187–202. https://doi.org/10.2307/2137845

MASON, A. and LEE, R. (2006). Reform and support systems for the elderly in developing countries: Capturing the second demographic dividend. *Genus* **62** 11–35.

O'NEILL, B. C., DALTON, M., FUCHS, R., JIANG, L. and PACHAURI, S. (2010). Global demographic trends and future carbon emissions. *Proc. Natl. Acad. Sci.* **107** 17521–17526. https://doi.org/10.1073/pnas.10045811

PEARL, J. (1993). Comment: Graphical models, causality and intervention. *Statist. Sci.* **8** 266–269. https://doi.org/10.1214/ss/1177010894

PRESTON, S. H., HEUVELINE, P. and GUILLOT, M., eds. (2001). *Demography: Measuring and Modeling Population Processes* Blackwell Publishers, Oxford.

RAFTERY, A. E., ALKEMA, L. and GERLAND, P. (2014). Bayesian population projections for the United Nations. *Statist. Sci.* **29** 58–68. MR3201847 https://doi.org/10.1214/13-STS419

RAFTERY, A. E., ZIMMER, A., FRIERSON, D. M. W., STARTZ, R. and LIU, P. (2017). Less than 2 C warming by 2100 unlikely. *Nat. Clim. Change* **7** 637–641. https://doi.org/10.1038/nclimate3352

ŠEVČÍKOVÁ, H., ALKEMA, L. and RAFTERY, A. E. (2011). bayesTFR: An R package for probabilistic projections of the total fertility rate. *J. Stat. Softw.* **43** 1–29. https://doi.org/10.18637/jss.v043.i01

ŠEVČÍKOVÁ, H. and RAFTERY, A. E. (2016). bayesPop: Probabilistic population projections. *J. Stat. Softw.* **75** 1–29. https://doi.org/10.18637/jss.v075.i05

SOBOTKA, T., BEAUJOUAN, É. and VAN BAVEL, J. (2017). Introduction: Education and fertility in low-fertility settings. *Vienna Yearb. Popul. Res.* **15** 1–16. https://doi.org/10.1553/populationyearbook2017s001

SPERINGER, M., GOUJON, A., KC, S., POTANČOKOVÁ, M., REITER, C., JURASSZOVICH, S. and EDER, J. (2019). Global reconstruction of educational attainment, 1950 to 2015: Methodology and assessment. Vienna Institute of Demography. Working Papers No. 02/2019. https://doi.org/10.1553/0x003cb434

UNITED NATIONS, DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS, POPULATION DIVISION (2019a). World population prospects: The 2019 revision. Online edition.

UNITED NATIONS, DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS, POPULATION DIVISION (2019b). World population prospects 2019: Methodology of the United Nations population estimates and projections.

UNITED NATIONS (2015). Transforming our world: The 2030 agenda for sustainable development. Available at https://sdgs.un.org/2030agenda.

VOLLSET, S. E., GOREN, E., YUAN, C.-W., CAO, J., SMITH, A. E., HSIAO, T., BISIGNANO, C., AZHAR, G. S., CASTRO, E. et al. (2020). Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: A forecasting analysis for the Global Burden of Disease study. *Lancet* **396** 1285–1306. https://doi.org/10.1016/S0140-6736(20)30677-2

WILMOTH, J. (2019). Global population projections: A critical comparison of key methods and assumptions. Available at https://www.un.org/en/development/desa/population/about/director/pdf/Wilmoth_APC_Nov2019_Script.pdf.

WITTGENSTEIN CENTRE FOR DEMOGRAPHY AND GLOBAL HUMAN CAPITAL (2018). Wittgenstein Centre data explorer version 2.0. Available at http://www.wittgensteincentre.org/dataexplorer.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 404–423
https://doi.org/10.1214/23-AOAS1794

A BAYESIAN SPATIO-TEMPORAL LEVEL SET DYNAMIC MODEL AND APPLICATION TO FIRE FRONT PROPAGATION

BY MYUNGSOO YOO AND CHRISTOPHER K. WIKLE[a]

Department of Statistics, University of Missouri, [a]*WikleC@Missouri.edu*

Intense wildfires impact nature, humans, and society, causing catastrophic damage to property and the ecosystem as well as the loss of life. Forecasting wildfire front propagation and understanding the behavior of wildfire propagation within a formal uncertainty quantification framework are essential in order to support fire fighting efforts and plan evacuations. The level set method has been widely used to analyze the change in surfaces, shapes, and boundaries. In particular, a signed distance function used in level set methods can readily be interpreted to represent complicated boundaries and their changes in time. While there is substantial literature on the level set method in wildfire applications, these implementations have relied on a heavily-parameterized formula for the rate of spread. These implementations have not typically considered uncertainty quantification, incorporated data-driven learning, nor summarized the effect of the environmental covariates. Here we present a Bayesian spatio-temporal dynamic model, based on level sets, which can be utilized for inference and forecasting the boundary of interest in the presence of uncertain data and lack of knowledge about the boundary velocity. The methodology relies on both a mechanistically-motivated dynamic model for level sets and a stochastic spatio-temporal dynamic model for the front velocity. We show the effectiveness of our method via simulation and with forecasting the fire front boundary evolution of two classic California megafires—the 2017–2018 Thomas fire and the 2017 Haypress fire.

REFERENCES

ADDISON, P. and OOMMEN, T. (2020). Post-fire debris flow modeling analyses: Case study of the post-Thomas fire event in California. *Nat. Hazards* **100** 329–343. https://doi.org/10.1007/s11069-019-03814-x

AGEE, J. K. (1996). The influence of forest structure on fire behavior. In *The* 17*th Annual Forest Vegetation Management Conference*.

ALESSANDRI, A., BAGNERINI, P., GAGGERO, M. and MANTELLI, L. (2021). Parameter estimation of fire propagation models using level set methods. *Appl. Math. Model.* **92** 731–747. MR4186583 https://doi.org/10.1016/j.apm.2020.11.030

BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. MR3362184

BROTONS, L., AQUILUÉ, N., DE CÁCERES, M., FORTIN, M.-J. and FALL, A. (2013). How fire history, fire suppression practices and climate change affect wildfire regimes in Mediterranean landscapes. *PLoS ONE* **8** e62392. https://doi.org/10.1371/journal.pone.0062392

CLARK, T. L., COEN, J. and LATHAM, D. (2004). Description of a coupled atmosphere–fire model. *Int. J. Wildland Fire* **13** 49–63. https://doi.org/10.1071/WF03043

CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2848400

DABROWSKI, J. J., HUSTON, C., HILTON, J., MANGEON, S. and KUHNERT, P. (2022). Towards data assimilation in level-set wildfire models using Bayesian filtering. Available at arXiv:2206.08501. https://doi.org/10.48550/arXiv.2206.08501

DUNLOP, M. M., IGLESIAS, M. A. and STUART, A. M. (2017). Hierarchical Bayesian level set inversion. *Stat. Comput.* **27** 1555–1584. MR3687326 https://doi.org/10.1007/s11222-016-9704-8

FENDELL, F. E. and WOLFF, M. F. (2001). Wind-aided fire spread. In *Forest Fires*: *Behavior and Ecological Effects* 171–223. Academic Press, San Diego, CA. https://doi.org/10.1016/B978-012386660-8/50008-8

Key words and phrases. Wildfire, signed distance function, boundary, level set method.

FILIPPI, J. B., BOSSEUR, F., MARI, C., LAC, C., MOIGNE, P. L., CUENOT, B., VEYNANTE, D., CARIOLLE, D. and BALBI, J.-H. (2009). Coupled atmosphere-wildland fire modelling. *J. Adv. Model. Earth Syst.* **11**. https://doi.org/10.3894/JAMES.2009.1.11

FINNEY, M. A. (2004). FARSITE: Fire Area Simulator—model development and evaluation. Res. Pap. RMRS-RP-4, Revised 2004, Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

GIBOU, F., FEDKIW, R. and OSHER, S. (2018). A review of level-set methods and some recent applications. *J. Comput. Phys.* **353** 82–109. MR3723645 https://doi.org/10.1016/j.jcp.2017.10.006

GOSLEE, S. (2019). landsat: Radiometric and Topographic Correction of Satellite Imagery. R package version 1.1.0.

GREEN, M. E., KAISER, K. and SHENTON, N. (2020). Modeling Wildfire Perimeter Evolution using Deep Neural Networks. Available at arXiv:2009.03977. https://doi.org/10.48550/arXiv.2009.03977

GEOSPATIAL MULTI-AGENCY COORDINATION GROUP (2019). "Historic Perimeters Combined 2000-2018 GeoMAC". Geospatial Multi-Agency Coordination Group.

HIJMANS, R. J., ETTEN, J. V., SUMNER, M., CHENG, J., BASTON, D., BEVAN, A., BIVAND, R., BUSETTO, L., CANTY, M. et al. (2022). raster: Geographic Data Analysis and Modeling. R package version 3.5-29.

HILTON, J. E., SULLIVAN, A. L., SWEDOSH, W., SHARPLES, J. and THOMAS, C. (2018). Incorporating convective feedback in wildfire simulations using pyrogenic potential. *Environ. Model. Softw.* **107** 12–24. https://doi.org/10.1016/j.envsoft.2018.05.009

HOLLISTER, J., SHAH, T., ROBITAILLE, A. L., BECK, M. W. and JOHNSON, M. (2022). elevatr: Access Elevation Data from Various APIs. R package version 0.4.2.

HOOTEN, M. B. and WIKLE, C. K. (2008). A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian collared-dove. *Environ. Ecol. Stat.* **15** 59–70. MR2412679 https://doi.org/10.1007/s10651-007-0040-1

HU, X., SUN, Y. and NTAIMO, L. (2012). DEVS-FIRE: Design and application of formal discrete event wildfire spread and suppression models. *Simulation* **88** 259–279. https://doi.org/10.1177/0037549711414592

IGLESIAS, M. A., LU, Y. and STUART, A. (2016). A Bayesian level set method for geometric inverse problems. *Interfaces Free Bound.* **18** 181–217. MR3549006 https://doi.org/10.4171/IFB/362

LANDFIRE (2017a). Forest Canopy Cover, LANDFIRE 1.4.0. U.S. Department of the Interior, Geological Survey, and U.S. Department of Agriculture.

LANDFIRE (2017b). Existing Vegetation Cover, LANDFIRE 1.4.0. U.S. Department of the Interior, Geological Survey, and U.S. Department of Agriculture.

LO, S. (2012). *A Fire Simulation Model for Heterogeneous Environments Using the Level Set Method*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—The Claremont Graduate University. MR3094046

MALLET, V., KEYES, D. E. and FENDELL, F. E. (2009). Modeling wildland fire propagation with level set methods. *Comput. Math. Appl.* **57** 1089–1101. MR2508539 https://doi.org/10.1016/j.camwa.2008.10.089

OSHER, S. and FEDKIW, R. (2003). *Level Set Methods and Dynamic Implicit Surfaces*. *Applied Mathematical Sciences* **153**. Springer, New York. MR1939127 https://doi.org/10.1007/b98879

OSHER, S. and SETHIAN, J. A. (1988). Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys.* **79** 12–49. MR0965860 https://doi.org/10.1016/0021-9991(88)90002-2

PAUSAS, J. G. and KEELEY, J. E. (2019). Wildfires as an ecosystem service. *Front. Ecol. Environ.* **17** 289–295. https://doi.org/10.1002/fee.2044

RAVENZWAAIJ, D. V., CASSEY, P. and BROWN, S. D. (2018). A simple introduction to Markov chain Monte–Carlo sampling. *Psychon. Bull. Rev.* **25** 143–154. https://doi.org/10.3758/s13423-016-1015-8

ROCHOUX, M. C., DELMOTTE, B., CUENOT, B., RICCI, S. and TROUVÉ, A. (2013). Regional-scale simulations of wildland fire spread informed by real-time flame front observations. *Proc. Combust. Inst.* 2641–2647. https://doi.org/10.1016/j.proci.2012.06.090

ROTHERMEL, R. C. (1972). A mathematical model for predicting fire spread in wildland fuels. Ogden, UT: U.S. Department of Agriculture, Intermountain Forest and Range Experiment Station. Res. Pap. INT-115.

SRIVAS, T., ARTÉS, T., DE CALLAFON, R. A. and ALTINTAS, I. (2016). Wildfire spread prediction and assimilation for FARSITE using ensemble Kalman filtering. *Proc. Comput. Sci.* **80** 897–908. https://doi.org/10.1016/j.procs.2016.05.328

VATS, D. and KNUDSON, C. (2021). Revisiting the Gelman–Rubin diagnostic. *Statist. Sci.* **36** 518–529. MR4323050 https://doi.org/10.1214/20-sts812

WAN, X., LI, C. and PARIKH, S. J. (2021). Chemical composition of soil-associated ash from the southern California Thomas fire and its potential inhalation risks to farmworkers. *J. Environ. Manag.* https://doi.org/10.1016/j.jenvman.2020.111570

WIKLE, C. K. (2010). Low-rank representations for spatial processes. In *Handbook of Spatial Statistics*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 107–118. CRC Press, Boca Raton, FL. MR2730946 https://doi.org/10.1201/9781420072884-c8

WIKLE, C. K. and HOOTEN, M. B. (2010). A general science-based framework for dynamical spatio-temporal models. *TEST* **19** 417–451. MR2745992 https://doi.org/10.1007/s11749-010-0209-z

WIKLE, C. K., ZAMMIT-MANGION, A. and CRESSIE, N. (2019). *Spatio-Temporal Statistics with R*. CRC Press/CRC, Florida.

XIE, J., EFENDIEV, Y. and DATTA-GUPTA, A. (2011). Uncertainty quantification in history matching of channelized reservoirs using Markov chain level set approaches. In 2011 *SPE Reservoir Simulation Symposium*. https://doi.org/10.2118/141811-MS

XUE, H., GU, F. and HU, X. (2012). Data assimilation using sequential Monte Carlo methods in wildfire spread simulation. *ACM Trans*. *Model*. *Comput*. *Simul*. 1–25. https://doi.org/10.1145/2379810.2379816

YOO, M. and WIKLE, C. K. (2023). Using echo state networks to inform physical models for fire front propagation. *Spat*. *Stat*. **54** Paper No. 100732, 17. MR4554665 https://doi.org/10.1016/j.spasta.2023.100732

YOO, M. and WIKLE, C. K. (2024). Supplement to "A Bayesian spatio-temporal level set dynamic model and application to fire front propagation." https://doi.org/10.1214/23-AOAS1794SUPPA, https://doi.org/10.1214/23-AOAS1794SUPPB

ZAMMIT-MANGION, A. and WIKLE, C. K. (2020). Deep integro-difference equation models for spatio-temporal forecasting. *Spat*. *Stat*. **37** 100408, 20. MR4109596 https://doi.org/10.1016/j.spasta.2020.100408

ZIMMERMAN, D. L., TANG, J. and HUANG, R. (2019). Outline analyses of the called strike zone in Major League Baseball. *Ann*. *Appl*. *Stat*. **13** 2416–2451. MR4037436 https://doi.org/10.1214/19-aoas1285

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 424–444
https://doi.org/10.1214/23-AOAS1795

NETWORK-LEVEL TRAFFIC FLOW PREDICTION: FUNCTIONAL TIME SERIES VS. FUNCTIONAL NEURAL NETWORK APPROACH

BY TAO MA[1,a], FANG YAO[2,b] AND ZHOU ZHOU[3,c]

[1]*Ingram School of Engineering, Texas State University,* [a]*tao.ma@txstate.edu*
[2]*Department of Probability & Statistics, School of Mathematical Sciences, Centre for Statistical Science, Peking University,* [b]*fyao@math.pku.edu.cn*
[3]*Department of Statistical Sciences, University of Toronto,* [c]*zhou@utstat.utoronto.ca*

Traffic state prediction is an essential component and an underlying backbone of intelligent transportation systems, especially in the context of smart city framework. Its significance is mainly twofold in modern transportation systems: supporting advanced traffic operations and management for highways and urban road networks to mitigate traffic congestion and enabling individual drivers with connected vehicles in the traffic system to dynamically optimize their routes to improve travel time. Traffic state prediction with interval-based pointwise methods at 15-minute or hourly intervals is common in traffic literature. However, because traffic dynamics are a continuous process over time, the discrete-time pointwise methods for traffic prediction at a fixed time interval hardly meet the advanced demands of continuous prediction in modern transportation systems. To close the gap, we propose functional approaches to intraday and day-by-day continuous-time prediction for traffic volume. This research focuses on network-level traffic flow predictions concurrently for all locations of interest. Two functional approaches are introduced, namely, the network-integrated functional time-series model and the functional neural network model. With functional approaches a 24-hour intraday traffic profile is modeled as a functional curve over time, and sequences of historical traffic curves are used to predict traffic curves for near future days in a row and multiple locations of interest. We also include the functional varying coefficient model, Sparse VAR and traditional AR models in the comparative study; empirical results show that the network-integrated functional time-series model outperforms other approaches in terms of the accuracy of predictions at network-scale.

REFERENCES

ABADI, A., RAJABIOUN, T. and IOANNOU, P. A. (2015). Traffic flow prediction for road transportation networks with limited traffic data. *IEEE Trans. Intell. Transp. Syst.* **16** 653–662.

ABDULHAI, B., PORWAL, H. and RECKER, W. (2002). Short-term traffic flow prediction using neuro-genetic algorithms. *J. Intell. Transp. Syst.* **7** 3–41.

ANEIROS, G., BONGIORNO, E. G., CAO, R. and VIEU, P., eds. (2017) *Functional Statistics and Related Fields. Contributions to Statistics*. Springer, Cham. MR3837566

ANEIROS-PÉREZ, G. and VIEU, P. (2006). Semi-functional partial linear regression. *Statist. Probab. Lett.* **76** 1102–1110. MR2269280 https://doi.org/10.1016/j.spl.2005.12.007

AUE, A., NORINHO, D. D. and HÖRMANN, S. (2015). On the prediction of stationary functional time series. *J. Amer. Statist. Assoc.* **110** 378–392. MR3338510 https://doi.org/10.1080/01621459.2014.909317

BASU, S. and MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. MR3357870 https://doi.org/10.1214/15-AOS1315

BESSE, P. C. and CARDOT, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canad. J. Statist.* **24** 467–487. MR1438418 https://doi.org/10.2307/3315328

BOSQ, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. In *Nonparametric Functional Estimation and Related Topics* (*Spetses*, 1990). *NATO Adv. Sci. Inst. Ser. C*: *Math. Phys. Sci.* **335** 509–529. Kluwer Academic, Dordrecht. MR1154349

Key words and phrases. Matrix-variate factor model, functional principal component, network-level traffic flow prediction, functional time series, functional neural network, functional varying coefficient.

BOSQ, D. (2000). *Linear Processes in Function Spaces*: *Theory and Applications*. *Lecture Notes in Statistics* **149**. Springer, New York. MR1783138 https://doi.org/10.1007/978-1-4612-1154-9

BROCKWELL, P. J. and DAVIS, R. A. (2016). *Introduction to Time Series and Forecasting*, 3rd ed. *Springer Texts in Statistics*. Springer, Cham. MR3526245 https://doi.org/10.1007/978-3-319-29854-2

CETIN, M. and COMERT, G. (2006). Short-term traffic flow prediction with regime switching models. *Transp*. *Res*. *Rec*. **1965** 23–31.

CHIOU, J.-M. (2012). Dynamical functional prediction and classification, with application to traffic flow prediction. *Ann*. *Appl*. *Stat*. **6** 1588–1614. MR3058676 https://doi.org/10.1214/12-AOAS595

CHIOU, J.-M. and LI, P.-L. (2007). Functional clustering and identifying substructures of longitudinal data. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **69** 679–699. MR2370075 https://doi.org/10.1111/j.1467-9868.2007.00605.x

CHIOU, J.-M. and MÜLLER, H.-G. (2007). Diagnostics for functional regression via residual processes. *Comput*. *Statist*. *Data Anal*. **51** 4849–4863. MR2364544 https://doi.org/10.1016/j.csda.2006.07.042

CHIOU, J.-M., MÜLLER, H.-G. and WANG, J.-L. (2003). Functional quasi-likelihood regression models with smooth random effects. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **65** 405–423. MR1983755 https://doi.org/10.1111/1467-9868.00393

CRAVEN, P. WAHBA, G. (1978). Smoothing noisy data with spline functions. *Numer*. *Math*. **31** 377–403.

CRAWFORD, F., WATLING, D. P. and CONNORS, R. D. (2017). A statistical method for estimating predictable differences between daily traffic flow profiles. *Transp*. *Res*., *Part B*, *Methodol*. **95** 196–213.

CYBENKO, G. (1989). Approximation by superpositions of a sigmoidal function. *Math*. *Control Signals Systems* **2** 303–314. MR1015670 https://doi.org/10.1007/BF02551274

DAI, X., FU, R., ZHAO, E., ZHANG, Z., LIN, Y., WANG, F.-Y. and LI, L. (2019). DeepTrend 2.0: A light-weighted multi-scale traffic prediction model using detrending. *Transp*. *Res*., *Part C*, *Emerg*. *Technol*. **103** 142–157.

DELL'ACQUA, P., BELLOTTI, F., BERTA, R. and DE GLORIA, A. (2015). Time-aware multivariate nearest neighbor regression methods for traffic flow prediction. *IEEE Trans*. *Intell*. *Transp*. *Syst*. **16** 3393–3402.

DENG, W., LEI, H. and ZHOU, X. (2013). Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach. *Transp*. *Res*., *Part B*, *Methodol*. **57** 132–157.

EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann*. *Statist*. **32** 407–499. MR2060166 https://doi.org/10.1214/009053604000000067

ERMAGUN, A. and LEVINSON, D. (2018). Spatiotemporal traffic forecasting: Review and proposed directions. *Transp*. *Rev*. **38** 786–814.

FENG, X., LING, X., ZHENG, H., CHEN, Z. and XU, Y. (2019). Adaptive multi-kernel SVM with spatial–temporal correlation for short-term traffic flow prediction. *IEEE Trans*. *Intell*. *Transp*. *Syst*. **20** 2001–2013.

FERRATY, F., ed. (2011) *Recent Advances in Functional Data Analysis and Related Topics*. *Contributions to Statistics*. Physica-Verlag/Springer, Heidelberg. MR2867568 https://doi.org/10.1007/978-3-7908-2736-1

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis*: *Theory and Practice*. *Springer Series in Statistics*. Springer, New York. MR2229687

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

GU, Y., LU, W., XU, X., QIN, L., SHAO, Z. and ZHANG, H. (2020). An improved Bayesian combination model for short-term traffic prediction with deep learning. *IEEE Trans*. *Intell*. *Transp*. *Syst*. **21** 1332–1342.

GUARDIOLA, I. G., LEON, T. and MALLOR, F. (2014). A functional approach to monitor and recognize patterns of daily traffic profiles. *Transp*. *Res*., *Part B*, *Methodol*. **65** 119–136.

GUO, X. and ZHU, Q. (2009). A traffic flow forecasting model based on BP neural network. In 2009 2*nd International Conference on Power Electronics and Intelligent Transportation System* (*PEITS*) **3** 311–314.

HÖRMANN, S. and KIDZIŃSKI, Ł. (2015). A note on estimation in Hilbertian linear models. *Scand*. *J*. *Stat*. **42** 43–62. MR3318024 https://doi.org/10.1111/sjos.12094

HÖRMANN, S., KIDZIŃSKI, Ł. and HALLIN, M. (2015). Dynamic functional principal components. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **77** 319–348. MR3310529 https://doi.org/10.1111/rssb.12076

HÖRMANN, S. and KOKOSZKA, P. (2010). Weakly dependent functional data. *Ann*. *Statist*. **38** 1845–1884. MR2662361 https://doi.org/10.1214/09-AOS768

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. Springer, New York. MR2920735 https://doi.org/10.1007/978-1-4614-3655-3

HORVÁTH, L., KOKOSZKA, P. and RICE, G. (2014). Testing stationarity of functional time series. *J*. *Econometrics* **179** 66–82. MR3153649 https://doi.org/10.1016/j.jeconom.2013.11.002

JIA, T. and YAN, P. (2020). Predicting citywide road traffic flow using deep spatiotemporal neural networks. *IEEE Trans*. *Intell*. *Transp*. *Syst*. 1–11.

KAMARIANAKIS, Y. and PRASTACOS, P. (2005). Space–time modeling of traffic flow. *Comput*. *Geosci*. **31** 119–133.

KARLAFTIS, M. G. and VLAHOGIANNI, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transp. Res., Part C, Emerg. Technol.* **19** 387–399.

KLEPSCH, J., KLÜPPELBERG, C. and WEI, T. (2017). Prediction of functional ARMA processes with an application to traffic data. *Econom. Stat.* **1** 128–149. MR3669993 https://doi.org/10.1016/j.ecosta.2016.10.009

LAHAROTTE, P.-A. (2016). Contributions à la prévision court-terme, multi-échelle et multi-variée, par apprentissage statistique du trafic routier Ph.D. Thesis, Université de Lyon.

LANA, I., DEL SER, J., VELEZ, M. and VLAHOGIANNI, E. I. (2018). Road traffic forecasting: Recent advances and new challenges. *IEEE Intell. Transp. Syst. Mag.* **10** 93–109.

LIU, L., ZHEN, J., LI, G., ZHAN, G., HE, Z., DU, B. and LIN, L. (2020). Dynamic spatial-temporal representation learning for traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* 1–15.

LV, Y., DUAN, Y., KANG, W., LI, Z. and WANG, F.-Y. (2015). Traffic flow prediction with big data: A deep learning approach. *IEEE Trans. Intell. Transp. Syst.* **16** 865–873.

MA, T., ANTONIOU, C. and TOLEDO, T. (2020). Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transp. Res., Part C, Emerg. Technol.* **111** 352–372.

MA, T., YAO, F. and ZHOU, Z. (2024). Supplement to "Network-level traffic flow prediction: Functional time series vs. functional neural network approach." https://doi.org/10.1214/23-AOAS1795SUPP

MA, T., ZHOU, Z. and ANTONIOU, C. (2018). Dynamic factor model for network traffic state forecast. *Transp. Res., Part B, Methodol.* **118** 281–317.

MA, X., DAI, Z., HE, Z., MA, J., WANG, Y. and WANG, Y. (2017). Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors* **17**. https://doi.org/10.3390/s17040818

MALLAT, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **11** 674–693.

MIN, W. and WYNTER, L. (2011). Real-time road traffic prediction with spatio-temporal correlations. *Transp. Res., Part C, Emerg. Technol.* **19** 606–616.

MITROVIC, N., ASIF, M. T., DAUWELS, J. and JAILLET, P. (2015). Low-dimensional models for compressed sensing and prediction of large-scale traffic data. *IEEE Trans. Intell. Transp. Syst.* **16** 2949–2954.

MÜLLER, H.-G. (2005). Functional modelling and classification of longitudinal data. *Scand. J. Stat.* **32** 223–246. MR2188671 https://doi.org/10.1111/j.1467-9469.2005.00429.x

MÜLLER, H.-G. (2009). Functional modeling of longitudinal data. In *Longitudinal Data Analysis. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 223–251. CRC Press, Boca Raton, FL. MR1500120

MÜLLER, H.-G. and ZHANG, Y. (2005). Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. *Biometrics* **61** 1064–1075. MR2216200 https://doi.org/10.1111/j.1541-0420.2005.00378.x

MUNOZ, L., SUN, X., HOROWITZ, R. and ALVAREZ, L. (2003). Traffic density estimation with the cell transmission model. In *Proceedings of the* 2003 *American Control Conference* **5** 3750–3755.

NANTHAWICHIT, C., NAKATSUJI, T. and SUZUKI, H. (2003). Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. *Transp. Res. Rec.* **1855** 49–59.

OH, S., KIM, Y. and HONG, J. (2015). Urban traffic flow prediction system using a multifactor pattern recognition model. *IEEE Trans. Intell. Transp. Syst.* **16** 2744–2755.

PERCIVAL, D. B. and WALDEN, A. T. (2000). *Wavelet Methods for Time Series Analysis. Cambridge Series in Statistical and Probabilistic Mathematics* **4**. Cambridge Univ. Press, Cambridge. MR1770693 https://doi.org/10.1017/CBO9780511841040

QIAO, F., YANG, H. and LAM, W. H. K. (2001). Intelligent simulation and prediction of traffic flow dispersion. *Transp. Res., Part B, Methodol.* **35** 843–863.

RAMSAY, J. O. and DALZELL, C. J. (1991). Some tools for functional data analysis. *J. Roy. Statist. Soc. Ser. B* **53** 539–572. MR1125714

RAMSAY, J. O. and SILVERMAN, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies. Springer Series in Statistics*. Springer, New York. MR1910407 https://doi.org/10.1007/b98886

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

RICE, J. A. (2004). Functional and longitudinal data analysis: Perspectives on smoothing. *Statist. Sinica* **14** 631–647. MR2087966

ROSSI, F. and CONAN-GUEZ, B. (2005). Functional multi-layer perceptron: A non-linear tool for functional data analysis. *Neural Netw.* **18** 45–60. https://doi.org/10.1016/j.neunet.2004.07.001

ROSSI, F., DELANNAY, N., CONAN-GUEZ, B. and VERLEYSEN, M. (2005). Representation of functional data in neural networks. *Neurocomputing* **64** 183–210.

TAN, H., WU, Y., SHEN, B., JIN, P. J. and RAN, B. (2016). Short-term traffic prediction based on dynamic tensor completion. *IEEE Trans. Intell. Transp. Syst.* **17** 2123–2133.

THIND, B., MULTANI, K. and CAO, J. (2023). Deep learning with functional inputs. *J. Comput. Graph. Statist.* **32** 171–180. MR4552945 https://doi.org/10.1080/10618600.2022.2097914

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

VLAHOGIANNI, E. I., KARLAFTIS, M. G. and GOLIAS, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transp. Res., Part C, Emerg. Technol.* **43** 3–19.

WAGNER-MUNS, I. M., GUARDIOLA, I. G., SAMARANAYKE, V. A. and KAYANI, W. I. (2018). A functional data analysis approach to traffic volume forecasting. *IEEE Trans. Intell. Transp. Syst.* **19** 878–888.

WANG, D., LIU, X. and CHEN, R. (2019). Factor models for matrix-valued high-dimensional time series. *J. Econometrics* **208** 231–248. MR3906969 https://doi.org/10.1016/j.jeconom.2018.09.013

WANG, J., CHEN, R. and HE, Z. (2019). Traffic speed prediction for urban transportation network: A path based deep learning approach. *Transp. Res., Part C, Emerg. Technol.* **100** 372–385.

WANG, J., DENG, W. and GUO, Y. (2014). New Bayesian combination method for short-term traffic flow forecasting. *Transp. Res., Part C, Emerg. Technol.* **43** 79–94.

WANG, Y., COPPOLA, P., TZIMITSI, A., MESSMER, A., PAPAGEORGIOU, M. and NUZZOLO, A. (2011). Real-time freeway network traffic surveillance: Large-scale field-testing results in southern Italy. *IEEE Trans. Intell. Transp. Syst.* **12** 548–562.

WANG, Y. and PAPAGEORGIOU, M. (2005). Real-time freeway traffic state estimation based on extended Kalman filter: A general approach. *Transp. Res., Part B, Methodol.* **39** 141–167.

WILLIAMS, B. M. and HOEL, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *J. Transp. Eng.* **129** 664–672.

YAO, F., LEI, E. and WU, Y. (2015). Effective dimension reduction for sparse functional data. *Biometrika* **102** 421–437. MR3371014 https://doi.org/10.1093/biomet/asv006

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 https://doi.org/10.1198/016214504000001745

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106 https://doi.org/10.1214/009053605000000660

YU, H., WU, Z., WANG, S., WANG, Y. and MA, X. (2017). Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors* **17**. https://doi.org/10.3390/s17071501

ZHANG, Y., ZHANG, Y. and HAGHANI, A. (2014). A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. *Transp. Res., Part C, Emerg. Technol.* **43** 65–78.

ZHAO, X., MARRON, J. S. and WELLS, M. T. (2004). The functional data analysis view of longitudinal data. *Statist. Sinica* **14** 789–808. MR2087973

ZHENG, W., LEE, D. and SHI, Q. (2006). Short-term freeway traffic flow prediction: Bayesian combined neural network approach. *J. Transp. Eng.* **132** 114–121.

ZHU, H., LI, R. and KONG, L. (2012). Multivariate varying coefficient model for functional responses. *Ann. Statist.* **40** 2634–2666. MR3097615 https://doi.org/10.1214/12-AOS1045

ZHU, S., ZHAO, Y., ZHANG, Y., LI, Q., WANG, W. and YANG, S. (2020). Short-term traffic flow prediction with wavelet and multi-dimensional Taylor network model. *IEEE Trans. Intell. Transp. Syst.* 1–6.

ZHU, X., CAI, Z. and MA, Y. (2022). Network functional varying coefficient model. *J. Amer. Statist. Assoc.* **117** 2074–2085. MR4528490 https://doi.org/10.1080/01621459.2021.1901718

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 445–467
https://doi.org/10.1214/23-AOAS1796

SELECTIVE INFERENCE FOR SPARSE MULTITASK REGRESSION WITH APPLICATIONS IN NEUROIMAGING

BY SNIGDHA PANIGRAHI[1,a], NATASHA STEWART[1,b], CHANDRA SRIPADA[2,d] AND ELIZAVETA LEVINA[1,c]

[1]*Department of Statistics, University of Michigan,* [a]*psnigdha@umich.edu,* [b]*nstew@umich.edu,* [c]*elevina@umich.edu*
[2]*Department of Philosophy, University of Michigan,* [d]*sripada@umich.edu*

Multitask learning is frequently used to model a set of related response variables from the same set of features, improving predictive performance and modeling accuracy relative to methods that handle each response variable separately. Despite the potential of multitask learning to yield more powerful inference than single-task alternatives, prior work in this area has largely omitted uncertainty quantification. Our focus in this paper is a common multitask problem in neuroimaging, where the goal is to understand the relationship between multiple cognitive task scores (or other subject-level assessments) and brain connectome data collected from imaging. We propose a framework for selective inference to address this problem, with the flexibility to: (i) jointly identify the relevant predictors for each task through a sparsity-inducing penalty and (ii) conduct valid inference in a model based on the estimated sparsity structure. Our framework offers a new conditional procedure for inference, based on a refinement of the selection event that yields a tractable selection-adjusted likelihood. This gives an approximate system of estimating equations for maximum likelihood inference, solvable via a single convex optimization problem, and enables us to efficiently form confidence intervals with approximately the correct coverage. Applied to both simulated data and data from the Adolescent Brain Cognitive Development (ABCD) study, our selective inference methods yield tighter confidence intervals than commonly used alternatives, such as data splitting. We also demonstrate through simulations that multitask learning with selective inference can more accurately recover true signals than single-task methods.

REFERENCES

ADELI, E., MENG, Y., LI, G., LIN, W. and SHEN, D. (2019). Multi-task prediction of infant cognitive scores from longitudinal incomplete neuroimaging data. *NeuroImage* **185** 783–792.

ANDERSON, E. D. and BARBEY, A. K. (2023). Investigating cognitive neuroscience theories of human intelligence: A connectome-based predictive modeling approach. *Hum. Brain Mapp.* **44** 1647–1665. https://doi.org/10.1002/hbm.26164

ANDREASEN, N. C., NOPOULOS, P., O'LEARY, D. S., MILLER, D. D., WASSINK, T. and FLAUM, M. (1999). Defining the phenotype of schizophrenia: Cognitive dysmetria and its neural mechanisms. *Biol. Psychiatry* **46** 908–920.

ANDREWS-HANNA, J. R., SMALLWOOD, J. and SPRENG, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Ann. N.Y. Acad. Sci.* **1316** 29–52. https://doi.org/10.1111/nyas.12360

BASTIAN, A. J. (2006). Learning to predict the future: The cerebellum adapts feedforward movement control. *Curr. Opin. Neurobiol.* **16** 645–649. https://doi.org/10.1016/j.conb.2006.08.016

BI, J., XIONG, T., YU, S., DUNDAR, M. and RAO, R. B. (2021). An improved multi-task learning approach with applications in medical diagnosis. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 117–132.

BINDER, J. R. and DESAI, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn. Sci.* **15** 527–536.

Key words and phrases. Multitask learning, multilevel lasso, joint sparsity, postselection inference, selective inference, neuroimaging, fMRI data.

BINDER, J. R., DESAI, R. H., GRAVES, W. W. and CONANT, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* **19** 2767–2796.

BLAIR, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behav. Brain Sci.* **29** 109–125.

CARROLL, J. B. et al. (1993). *Human Cognitive Abilities*: *A Survey of Factor-Analytic Studies* 1. Cambridge Univ. Press, Cambridge.

CATTELL, R. B. (1943). The measurement of adult intelligence. *Psychol. Bull.* **40** 153.

CHEN, J., TAM, A., KEBETS, V., ORBAN, C., OOI, L. Q. R., ASPLUND, C. L., MAREK, S., DOSENBACH, N. U. F., EICKHOFF, S. B., BZDOK, D. et al. (2022). Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study. *Nat. Commun.* **13** 2217.

CHEN, Y., SPAGNA, A., WU, T., KIM, T. H., WU, Q., CHEN, C., WU, Y. and FAN, J. (2019). Testing a cognitive control model of human intelligence. *Sci. Rep.* **9** 1–17.

COLE, M. W., REYNOLDS, J. R., POWER, J. D., REPOVS, G., ANTICEVIC, A. and BRAVER, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* **16** 1348–1355. https://doi.org/10.1038/nn.3470

COLE, M. W. and SCHNEIDER, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage* **37** 343–360. https://doi.org/10.1016/j.neuroimage.2007.03.071

CORDES, D. and NANDY, R. R. (2006). Estimation of the intrinsic dimensionality of fMRI data. *NeuroImage* **29** 145–154. https://doi.org/10.1016/j.neuroimage.2005.07.054

COX, D. R. (1975). A note on data-splitting for the evaluation of significance levels. *Biometrika* **62** 441–444. MR0378189 https://doi.org/10.1093/biomet/62.2.441

FINN, E. S., SHEN, X., SCHEINOST, D., ROSENBERG, M. D., HUANG, J., CHUN, M. M., PAPADEMETRIS, X. and CONSTABLE, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18** 1664–1671. https://doi.org/10.1038/nn.4135

FITHIAN, W., SUN, D. and TAYLOR, J. (2017). Optimal inference after model selection. Available at arXiv:1410.2597.

FRAENZ, C., SCHLÜTER, C., FRIEDRICH, P., JUNG, R. E., GÜNTÜRKÜN, O. and GENÇ, E. (2021). Interindividual differences in matrix reasoning are linked to functional connectivity between brain regions nominated by parieto-frontal integration theory. *Intelligence* **87** 101545.

GIGNAC, G. E. (2014). Dynamic mutualism versus g factor theory: An empirical test. *Intelligence* **42** 89–97.

GORDON, E. M., LAUMANN, T. O., ADEYEMO, B., GILMORE, A. W., NELSON, S. M., DOSENBACH, N. U. F. and PETERSEN, S. E. (2017). Individual-specific features of brain systems identified with resting state functional correlations. *NeuroImage* **146** 918–939.

GORDON, E. M., LAUMANN, T. O., ADEYEMO, B., HUCKINS, J. F., KELLEY, W. M. and PETERSEN, S. E. (2016). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* **26** 288–303. https://doi.org/10.1093/cercor/bhu239

GRAY, J. R., CHABRIS, C. F. and BRAVER, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nat. Neurosci.* **6** 316–322. https://doi.org/10.1038/nn1014

GUO, J., LEVINA, E., MICHAILIDIS, G. and ZHU, J. (2011). Joint estimation of multiple graphical models. *Biometrika* **98** 1–15. MR2804206 https://doi.org/10.1093/biomet/asq060

HAGLER JR., D. J., HATTON, S., CORNEJO, M. D., MAKOWSKI, C., FAIR, D. A., DICK, A. S., SUTHERLAND, M. T., CASEY, B. J., BARCH, D. M. et al. (2019). Image processing and analysis methods for the adolescent brain cognitive development study. *NeuroImage* **202** 116091.

HEARNE, L. J., MATTINGLEY, J. B. and COCCHI, L. (2016). Functional brain networks related to individual differences in human intelligence at rest. *Sci. Rep.* **6** 1–8.

HONG, L., KUFFNER, T. A. and MARTIN, R. (2018). On overfitting and post-selection uncertainty assessments. *Biometrika* **105** 221–224. MR3768876 https://doi.org/10.1093/biomet/asx083

HORN, J. L. and NOLL, J. (1997). *Human cognitive capabilities*: *Gf-Ge theory*. Guilford, New York.

HÜLÜR, G., WILHELM, O. and ROBITZSCH, A. (2011). Intelligence differentiation in early childhood. *J. Individ. Differ.*

HUMPHREYS, L. G. (1979). The construct of general intelligence. *Intelligence* **3** 105–120.

JOLLIFFE, I. T. (2003). Principal component analysis. *Technometrics* **45** 276.

JUAN-ESPINOSA, M., GARCÍA LUÍS, F., COLOM, R. and ABAD, F. J. (2000). Testing the age related differentiation hypothesis through the Wechsler's scales. *Pers. Individ. Differ.* **29** 1069–1075.

KARCHER, N. R. and BARCH, D. M. (2021). The ABCD study: Understanding the development of risk for mental and physical health outcomes. *Neuropsychopharmacology* **46** 131–142. https://doi.org/10.1038/s41386-020-0736-6

KIVARANOVIC, D. and LEEB, H. (2018). Expected length of post-model-selection confidence intervals conditional on polyhedral constraints. Preprint. Available at arXiv:1803.01665.

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

LIU, K., MARKOVIC, J. and TIBSHIRANI, R. (2018). More powerful post-selection inference. with application to the Lasso. Preprint. Available at arXiv:1801.09037.

LOZANO, A. C. and SWIRSZCZ, G. (2012). Multi-level lasso for sparse multi-task regression. In *Proceedings of the* 29*th International Conference on Machine Learning* 595–602.

LUCIANA, M., BJORK, J. M., NAGEL, B. J., BARCH, D. M., GONZALEZ, R., NIXON, S. J. and BANICH, M. T. (2018). Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Dev. Cogn. Neurosci.* **32** 67–79. https://doi.org/10.1016/j.dcn.2018.02.006

MAREK, S., TERVO-CLEMMENS, B., CALABRO, F. J. MONTEZ, D. F., KAY, B. P., HATOUM, A. S., DONOHUE, M. R., FORAN, W., MILLER, R. L. et al. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature* **603** 654–660.

MARKETT, S., REUTER, M., HEEREN, B., LACHMANN, B., WEBER, B. and MONTAG, C. (2018). Working memory capacity and the functional connectome - insights from resting-state fMRI and voxelwise centrality mapping. *Brain Imaging Behav.* **12** 238–246. https://doi.org/10.1007/s11682-017-9688-9

NIENDAM, T. A., LAIRD, A. R., RAY, K. L., DEAN, Y. M., GLAHN, D. C. and CARTER, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn. Affect. Behav. Neurosci.* **12** 241–268.

PANIGRAHI, S. (2018). Carving model-free inference. Preprint. Available at arXiv:1811.03142.

PANIGRAHI, S., MOHAMMED, S., RAO, A. and BALADANDAYUTHAPANI, V. (2023). Integrative Bayesian models using post-selective inference: A case study in radiogenomics. *Biometrics* **79** 1801–1813. https://doi.org/10.1111/biom.13740

PANIGRAHI, S., STEWART, N., SRIPADA, C. and LEVINA, E. (2024). Supplement to "Selective inference for sparse multitask regression with applications in neuroimaging." https://doi.org/10.1214/23-AOAS1796SUPP

PANIGRAHI, S. and TAYLOR, J. (2018). Scalable methods for Bayesian selective inference. *Electron. J. Stat.* **12** 2355–2400. MR3832095 https://doi.org/10.1214/18-EJS1452

PANIGRAHI, S. and TAYLOR, J. (2022). Approximate selective inference via maximum likelihood. *J. Amer. Statist. Assoc.* Forthcoming.

PANIGRAHI, S., TAYLOR, J. and WEINSTEIN, A. (2021). Integrative methods for post-selection inference under convex constraints. *Ann. Statist.* **49** 2803–2824. MR4338384 https://doi.org/10.1214/21-aos2057

PANIGRAHI, S., WANG, J. and HE, X. (2022). Treatment Effect Estimation with Efficient Data Aggregation. Preprint. Available at arXiv:2203.12726.

SCHMAHMANN, J. D. (1996). From movement to thought: Anatomic substrates of the cerebellar contribution to cognitive processing. *Hum. Brain Mapp.* **4** 174–198.

SCHMAHMANN, J. D. (2019). The cerebellum and cognition. *Neurosci. Lett.* **688** 62–75. https://doi.org/10.1016/j.neulet.2018.07.005

SIMPSON-KENT, I. L., FUHRMANN, D., BATHELT, J., ACHTERBERG, J., BORGEEST, G. S. and KIEVIT, R. A. (2020). Neurocognitive reorganization between crystallized intelligence, fluid intelligence and white matter microstructure in two age-heterogeneous developmental cohorts. *Dev. Cogn. Neurosci.* **41** 100743.

SPEARMAN, C. (1961). "General Intelligence" Objectively Determined and Measured.

SPENCER, R. M. C., IVRY, R. B. and ZELAZNIK, H. N. (2005). Role of the cerebellum in movements: Control of timing or movement transitions? *Exp. Brain Res.* **161** 383–396. https://doi.org/10.1007/s00221-004-2088-6

SRIPADA, C., ANGSTADT, M., RUTHERFORD, S., KESSLER, D., KIM, Y., YEE, M. and LEVINA, E. (2019). Basic units of inter-individual variation in resting state connectomes. *Sci. Rep.* **9** 1–12.

SRIPADA, C., ANGSTADT, M., TAXALI, A., CLARK, D. A., GREATHOUSE, T., RUTHERFORD, S., DICKENS, J. R., SHEDDEN, K., GARD, A. M. et al. (2021). Brain-wide functional connectivity patterns support general cognitive ability and mediate effects of socioeconomic status in youth. *Transl. Psychiatry* **11** 1–8.

SRIPADA, C., RUTHERFORD, S., ANGSTADT, M., THOMPSON, W. K., LUCIANA, M., WEIGARD, A., HYDE, L. H. and HEITZEG, M. (2020). Prediction of neurocognition in youth from resting state fMRI. *Mol. Psychiatry* **25** 3413–3421. https://doi.org/10.1038/s41380-019-0481-6

SUZUMURA, S., NAKAGAWA, K., UMEZU, Y., TSUDA, K. and TAKEUCHI, I. (2017). Selective inference for sparse high-order interaction models. In *International Conference on Machine Learning* 3338–3347. PMLR.

TADAYON, E., PASCUAL-LEONE, A. and SANTARNECCHI, E. (2020). Differential contribution of cortical thickness, surface area, and gyrification to fluid and crystallized intelligence. *Cereb. Cortex* **30** 215–225. https://doi.org/10.1093/cercor/bhz082

TANIZAKI, K., HASHIMOTO, N., INATSU, Y., HONTANI, H. and TAKEUCHI, I. (2020). Computing valid p-values for image segmentation by selective inference. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 9553–9562.

TAYLOR, J. and TIBSHIRANI, R. (2018). Post-selection inference for ℓ_1-penalized likelihood models. *Canad. J. Statist.* **46** 41–61. MR3767165 https://doi.org/10.1002/cjs.11313

TIAN, X., PANIGRAHI, S. MARKOVIC, J., BI, N. and TAYLOR, J. (2016). Selective sampling after solving a convex problem. Preprint. Available at arXiv:1609.05609.

TIAN, X. and TAYLOR, J. (2018). Selective inference with a randomized response. *Ann. Statist.* **46** 679–710. MR3782381 https://doi.org/10.1214/17-AOS1564

TONG, X., XIE, H., CARLISLE, N., FONZO, G. A., OATHES, D. J., JIANG, J. and ZHANG, Y. (2022). Transdiagnostic connectome signatures from resting-state fMRI predict individual-level intellectual capacity. *Transl. Psychiatry* **12** 367.

VOLKOW, N. D., KOOB, G. F., CROYLE, R. T., BIANCHI, D. W., GORDON, J. A., KOROSHETZ, W. J., PÉREZ-STABLE, E. J., RILEY, W. T., BLOCH, M. H., CONWAY, K. et al. (2018). The conception of the ABCD study: From substance use to a broad NIH collaboration. *Dev. Cogn. Neurosci.* **32** 4–7.

WANG, J., HE, X. and XU, G. (2020). Debiased inference on treatment effect in a high-dimensional model. *J. Amer. Statist. Assoc.* **115** 442–454. MR4078474 https://doi.org/10.1080/01621459.2018.1558062

WANG, X., BI, J., YU, S., SUN, J. and SONG, M. (2016). Multiplicative multitask feature learning. *J. Mach. Learn. Res.* **17** Paper No. 80, 33. MR3517103

WIRTH, M., JANN, K., DIERKS, T., FEDERSPIEL, A., WIEST, R. and HORN, H. (2011). Semantic memory involvement in the default mode network: A functional neuroimaging study using independent component analysis. *NeuroImage* **54** 3057–3066. https://doi.org/10.1016/j.neuroimage.2010.10.039

ZHANG, Y. and YANG, Q. (2021). A survey on multi-task learning. *IEEE Trans. Knowl. Data Eng.*

ZHAO, Q. and PANIGRAHI, S. (2019). Selective inference for effect modification: An empirical investigation. *Obs. Stud.* **5** 131–140.

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443 https://doi.org/10.1214/009053607000000802

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 468–486
https://doi.org/10.1214/23-AOAS1797

LATENT SUBGROUP IDENTIFICATION IN IMAGE-ON-SCALAR REGRESSION

BY ZIKAI LIN[1,a], YAJUAN SI[2,c] AND JIAN KANG[1,b]

[1]*Department of Biostatistics, University of Michigan,* [a]*zikai@umich.edu,* [b]*jiankang@umich.edu*
[2]*Survey Research Center, Institute for Social Research, University of Michigan,* [c]*yajuan@umich.edu*

Image-on-scalar regression has been a popular approach to modeling the association between brain activities and scalar characteristics in neuroimaging research. The associations could be heterogeneous across individuals in the population, as indicated by recent large-scale neuroimaging studies, for example, the Adolescent Brain Cognitive Development (ABCD) Study. The ABCD data can inform our understanding of heterogeneous associations and how to leverage the heterogeneity and tailor interventions to increase the number of youths who benefit. It is of great interest to identify subgroups of individuals from the population such that: (1) within each subgroup the brain activities have homogeneous associations with the clinical measures; (2) across subgroups the associations are heterogeneous, and (3) the group allocation depends on individual characteristics. Existing image-on-scalar regression methods and clustering methods cannot directly achieve this goal. We propose a latent subgroup image-on-scalar regression model (LASIR) to analyze large-scale, multisite neuroimaging data with diverse sociodemographics. LASIR introduces the latent subgroup for each individual and group-specific, spatially varying effects, with an efficient stochastic expectation maximization algorithm for inferences. We demonstrate that LASIR outperforms existing alternatives for subgroup identification of brain activation patterns with functional magnetic resonance imaging data via comprehensive simulations and applications to the ABCD study. We have released our reproducible codes for public use with the software package available on Github.

REFERENCES

AKSHOOMOFF, N., BEAUMONT, J. L., BAUER, P. J., DIKMEN, S. S., GERSHON, R. C., MUNGAS, D., SLOTKIN, J., TULSKY, D., WEINTRAUB, S. et al. (2013). VIII. NIH Toolbox Cognition Battery (CB): Composite scores of crystallized, fluid, and overall cognition. *Monogr. Soc. Res. Child Dev.* **78** 119–132.

ASHBURNER, J. (2012). SPM: A history. *NeuroImage* **62** 791–800. https://doi.org/10.1016/j.neuroimage.2011.10.025

ASSARI, S., BOYCE, S. and JOVANOVIC, T. (2021). Association between hippocampal volume and working memory in 10,000+ 9-10-year-old children: Sex differences. *Child.* **8** 411.

ASSARI, S., BOYCE, S., SAQIB, M., BAZARGAN, M. and CALDWELL, C. H. (2021). Parental education and left lateral orbitofrontal cortical activity during N-back task: An fMRI study of American adolescents. *Brain Sci.* **11**. https://doi.org/10.3390/brainsci11030401

BRODERSEN, K. H., SCHOFIELD, T. M., LEFF, A. P., ONG, C. S., LOMAKINA, E. I., BUHMANN, J. M. and STEPHAN, K. E. (2011). Generative embedding for model-based classification of fMRI data. *PLoS Comput. Biol.* **7** e1002079, 19. MR2832099 https://doi.org/10.1371/journal.pcbi.1002079

CASEY, B., CANNONIER, T., CONLEY, M. I., COHEN, A. O., BARCH, D. M., HEITZEG, M. M., SOULES, M. E., TESLOVICH, T., DELLARCO, D. V. et al. (2018). The adolescent brain cognitive development (ABCD) study: Imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* **32** 43–54.

CLARK, U. S., MILLER, E. R. and HEGDE, R. R. (2018). Experiences of discrimination are associated with greater resting amygdala activity and functional connectivity. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3** 367–378.

Key words and phrases. Voxelwise spatial correlation, image-on-scalar regression, subgroup identification, stochastic expectation maximization.

COX, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29** 162–173. https://doi.org/10.1006/cbmr.1996.0014

DIEBOLT, J. and IP, E. H. S. (1996). Stochastic EM: Method and application. In *Markov Chain Monte Carlo in Practice. Interdiscip. Statist.* 259–273. CRC Press, London. MR1397971

EGLI, T., COYNEL, D., SPALEK, K., FASTENRATH, M., FREYTAG, V., HECK, A., LOOS, E., AUSCHRA, B., PAPASSOTIROPOULOS, A. et al. (2018). Identification of two distinct working memory-related brain networks in healthy young adults. *Eneuro* **5**.

EWING, S. W. F., BJORK, J. M. and LUCIANA, M. (2018). Implications of the ABCD study for developmental neuroscience. *Dev. Cogn. Neurosci.* **32** 161–164.

FALK, E. B., HYDE, L. W., MITCHELL, C., FAUL, J., GONZALEZ, R., HEITZEG, M. M., KEATING, D. P., LANGA, K. M., MARTZ, M. E. et al. (2013). What is a representative brain? Neuroscience meets population science. *Proc. Natl. Acad. Sci. USA* **110** 17615–17622.

GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

GROPPE, D. M., URBACH, T. P. and KUTAS, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiol.* **48** 1711–1725.

GRUN, B. and LEISCH, F. (2008). FlexMix version 2: Finite mixtures with concomitant variables and varying and constant parameters.

HAGLER JR, D. J., HATTON, S., CORNEJO, M. D., MAKOWSKI, C., FAIR, D. A., DICK, A. S., SUTHERLAND, M. T., CASEY, B., BARCH, D. M. et al. (2019). Image processing and analysis methods for the adolescent brain cognitive development study. *NeuroImage* **202** 116091.

HARTIGAN, J. A. and WONG, M. A. (1979). A K-Means clustering algorithm. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **28** 100–108.

HOLLAND, D., KUPERMAN, J. M. and DALE, A. M. (2010). Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo Planar Imaging. *NeuroImage* **50** 175–183. https://doi.org/10.1016/j.neuroimage.2009.11.044

JENKINSON, M., BECKMANN, C. F., BEHRENS, T. E. J., WOOLRICH, M. W. and SMITH, S. M. (2012). FSL. *NeuroImage* **62** 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

KANG, J. (2020). BayesGPfit: Fast Bayesian Gaussian process regression fitting. R package version 0.1.0.

KNOLL, L. J., OBLESER, J., SCHIPKE, C. S., FRIEDERICI, A. D. and BRAUER, J. (2012). Left prefrontal cortex activation during sentence comprehension covaries with grammatical knowledge in children. *NeuroImage* **62** 207–216. https://doi.org/10.1016/j.neuroimage.2012.05.014

LECEI, A., VAN HULST, B. M., DE ZEEUW, P., VAN DER PLUIJM, M., RIJKS, Y. and DURSTON, S. (2019). Can we use neuroimaging data to differentiate between subgroups of children with ADHD symptoms: A proof of concept study using latent class analysis of brain activity. *NeuroImage Clin.* **21** 101601. https://doi.org/10.1016/j.nicl.2018.11.011

LI, X., WANG, L. and WANG, H. J. (2021). Sparse learning and structure identification for ultrahigh-dimensional image-on-scalar regression. *J. Amer. Statist. Assoc.* **116** 1994–2008. Authors writing for the Alzheimer's Disease Neuroimaging Initiative. MR4353728 https://doi.org/10.1080/01621459.2020.1753523

LI, Y., ZHU, H., SHEN, D., LIN, W., GILMORE, J. H. and IBRAHIM, J. G. (2011). Multiscale adaptive regression models for neuroimaging data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 559–578. MR2853730 https://doi.org/10.1111/j.1467-9868.2010.00767.x

LIN, Z., SI, Y. and KANG, J. (2024). Supplement to "Latent subgroup identification in image-on-scalar regression." https://doi.org/10.1214/23-AOAS1797SUPP

LIU, Y., YAN, B., MERIKANGAS, K. and SHOU, H. (2017). Total variation regularized tensor-on-scalar regression. arXiv preprint. Available at arXiv:1703.05264.

MACDONALD, B., RANJAN, P. and CHIPMAN, H. (2015). GPfit: An R package for fitting a Gaussian process model to deterministic simulator outputs. *J. Stat. Softw.* **64** 1–23.

MAREK, S., TERVO-CLEMMENS, B., CALABRO, F. J. et al. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature* **603** 654–660.

MCKEOWN, M. J., HANSEN, L. K. and SEJNOWSK, T. J. (2003). Independent component analysis of functional MRI: What is signal and what is noise? *Curr. Opin. Neurobiol.* **13** 620–629.

MUSCHELLI, J., SWEENEY, E., LINDQUIST, M. and CRAINICEANU, C. (2015). fslr: Connecting the FSL software with R. *The R Journal* **7** 163.

NICHOLS, T. and HAYASAKA, S. (2003). Controlling the familywise error rate in functional neuroimaging: A comparative review. *Stat. Methods Med. Res.* **12** 419–446. MR2005445 https://doi.org/10.1191/0962280203sm341ra

PAUS, T. (2010). Population neuroscience: Why and how. *Hum. Brain Mapp.* **31** 891–903. https://doi.org/10.1002/hbm.21069

POWER, J. D., COHEN, A. L., NELSON, S. M., WIG, G. S., BARNES, K. A., CHURCH, J. A., VOGEL, A. C., LAUMANN, T. O., MIEZIN, F. M. et al. (2011). Functional network organization of the human brain. *Neuron* **72** 665–678.

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

ROUSSON, M., BROX, T. and DERICHE, R. (2003). Active unsupervised texture segmentation on a diffusion based feature space. In 2003 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003. *Proceedings*. **2** II–699. IEEE.

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

SHERMAN, L. E., RUDIE, J. D., PFEIFER, J. H., MASTEN, C. L., MCNEALY, K. and DAPRETTO, M. (2014). Development of the default mode and central executive networks across early adolescence: A longitudinal study. *Dev. Cogn. Neurosci.* **10** 148–159. https://doi.org/10.1016/j.dcn.2014.08.002

SHI, R. and KANG, J. (2015). Thresholded multiscale gaussian processes with application to Bayesian feature selection for massive neuroimaging data. arXiv preprint. Available at arXiv:1504.06074.

SI, Y., PALTA, M. and SMITH, M. (2020). Bayesian profiling multiple imputation for missing hemoglobin values in electronic health records. *Ann. Appl. Stat.* **14** 1903–1924. MR4194253 https://doi.org/10.1214/20-AOAS1378

SI, Y., REITER, J. P. and HILLYGUS, D. S. (2016). Bayesian latent pattern mixture models for handling attrition in panel studies with refreshment samples. *Ann. Appl. Stat.* **10** 118–143. MR3480490 https://doi.org/10.1214/15-AOAS876

SRIPADA, C., RUTHERFORD, S., ANGSTADT, M., THOMPSON, W. K., LUCIANA, M., WEIGARD, A., HYDE, L. H. and HEITZEG, M. (2020). Prediction of neurocognition in youth from resting state fMRI. *Mol. Psychiatry* **25** 3413–3421. https://doi.org/10.1038/s41380-019-0481-6

STERNBERG, R. J. and GRIGORENKO, E. L. (2002). *The General Factor of Intelligence*: *How General Is It?* Psychology Press, Mahwah, NJ, London.

WELLS III, W. M., VIOLA, P., ATSUMI, H., NAKAJIMA, S. and KIKINIS, R. (1996). Multi-modal volume registration by maximization of mutual information. *Med. Image Anal.* **1** 35–51.

WU, B., GUO, Y. and KANG, J. (2022). Bayesian spatial blind source separation via the thresholded Gaussian process. *J. Amer. Statist. Assoc.* 1–12.

ZHANG, D., LI, L., SRIPADA, C. and KANG, J. (2020). Image-on-scalar regression via deep neural networks. arXiv preprint. Available at arXiv:2006.09911.

ZHU, H., FAN, J. and KONG, L. (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *J. Amer. Statist. Assoc.* **109** 1084–1098. MR3265682 https://doi.org/10.1080/01621459.2014.881742

ZUO, X.-N., HE, Y., SU, X., HOU, X.-H., WENG, X. and LI, Q. (2018). Developmental population neuroscience: Emerging from ICHBD. *Sci. Bull.* **63** 331–332.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 487–505
https://doi.org/10.1214/23-AOAS1798

RETROSPECTIVE VARYING COEFFICIENT ASSOCIATION ANALYSIS OF LONGITUDINAL BINARY TRAITS: APPLICATION TO THE IDENTIFICATION OF GENETIC LOCI ASSOCIATED WITH HYPERTENSION

BY GANG XU[1,a], AMEI AMEI[1,b], WEIMIAO WU[2,d], YUNQING LIU[2,e],
LINCHUAN SHEN[1,c], EDWIN C. OH[3,g] AND ZUOHENG WANG[2,f]

[1]*Department of Mathematical Sciences, University of Nevada,* [a]*xug3@unlv.nevada.edu,* [b]*amei.amei@unlv.edu,* [c]*shenl4@unlv.nevada.edu*

[2]*Department of Biostatistics, Yale School of Public Health,* [d]*weimiao.wu@yale.edu,* [e]*yunqing.liu@yale.edu,* [f]*zuoheng.wang@yale.edu*

[3]*Department of Internal Medicine, University of Nevada School of Medicine,* [g]*edwin.oh@unlv.edu*

Many genetic studies contain rich information on longitudinal phenotypes that require powerful analytical tools for optimal analysis. Genetic analysis of longitudinal data that incorporates temporal variation is important for understanding the genetic architecture and biological variation of complex diseases. Most of the existing methods assume that the contribution of genetic variants is constant over time and fail to capture the dynamic pattern of disease progression. However, the relative influence of genetic variants on complex traits fluctuates over time. In this study, we propose a retrospective varying coefficient mixed model association test, RVMMAT, to detect time-varying genetic effect on longitudinal binary traits. We model dynamic genetic effect using smoothing splines, estimate model parameters by maximizing a double penalized quasi-likelihood function, design a joint test using a Cauchy combination method, and evaluate statistical significance via a retrospective approach to achieve robustness to model misspecification. Through simulations we illustrated that the retrospective varying-coefficient test was robust to model misspecification under different ascertainment schemes and gained power over the association methods assuming constant genetic effect. We applied RVMMAT to a genome-wide association analysis of longitudinal measure of hypertension in the Multi-Ethnic Study of Atherosclerosis. Pathway analysis identified two important pathways related to G-protein signaling and DNA damage. Our results demonstrated that RVMMAT could detect biologically relevant loci and pathways in a genome scan and provided insight into the genetic architecture of hypertension.

REFERENCES

BILD, D. E., BLUEMKE, D. A., BURKE, G. L., DETRANO, R., DIEZ ROUX, A. V., FOLSOM, A. R. et al. (2002). Multi-ethnic study of atherosclerosis: Objectives and design. *Amer. J. Epidemiol.* **156** 871–881.

BRESLOW, N. E. and CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88** 9–25.

BRESLOW, N. E. and LIN, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* **82** 81–91. MR1332840 https://doi.org/10.1093/biomet/82.1.81

BRYOIS, J., BUIL, A., FERREIRA, P. G., PANOUSIS, N. I., BROWN, A. A., VIÑUELA, A. et al. (2017). Time-dependent genetic effects on gene expression implicate aging processes. *Genome Res.* **27** 545–552.

CHEN, H., WANG, C., MATTHEW, A., LI, Z., SOFER, T. et al. (2016). Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am. J. Hum. Genet.* **98** 653–666.

CHEN, Z., BOEHNKE, M., WEN, X. and MUKHERJEE, B. (2021). Revisiting the genome-wide significance threshold for common variant GWAS. *G3* **11** jkaa056.

Key words and phrases. Genome-wide association study, varying coefficient model, generalized linear mixed model, time-varying genetic effect, model misspecification.

Chiang, C.-T., Rice, J. A. and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *J. Amer. Statist. Assoc.* **96** 605–619. MR1946428 https://doi.org/10.1198/016214501753168280

Chu, W., Li, R., Liu, J. and Reimherr, M. (2020). Feature selection for generalized varying coefficient mixed-effect models with application to obesity GWAS. *Ann. Appl. Stat.* **14** 276–298. MR4085094 https://doi.org/10.1214/19-AOAS1310

Chu, W., Li, R. and Reimherr, M. (2016). Feature screening for time-varying coefficient models with ultrahigh-dimensional longitudinal data. *Ann. Appl. Stat.* **10** 596–617. MR3528353 https://doi.org/10.1214/16-AOAS912

Das, K., Li, J., Wang, Z., Tong, C., Fu, G., Li, Y. et al. (2011). A dynamic model for genome-wide association studies. *Hum. Genet.* **129** 629–639.

Eubank, R. L., Huang, C., Muñoz Maldonado, Y., Wang, N., Wang, S. and Buchanan, R. J. (2004). Smoothing spline estimation in varying-coefficient models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 653–667. MR2088294 https://doi.org/10.1111/j.1467-9868.2004.B5595.x

Evangelou, E., Warren, H. R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H. et al. (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50** 1412–1425.

Fan, J., Ma, Y. and Dai, W. (2014). Nonparametric independence screening in sparse ultra-high-dimensional varying coefficient models. *J. Amer. Statist. Assoc.* **109** 1270–1284. MR3265696 https://doi.org/10.1080/01621459.2013.879828

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27** 1491–1518. MR1742497 https://doi.org/10.1214/aos/1017939139

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Stat. Interface* **1** 179–195. MR2425354 https://doi.org/10.4310/SII.2008.v1.n1.a15

Furlotte, N. A., Eskin, E. and Eyheramendy, S. (2012). Genome-wide association mapping with longitudinal data. *Genet. Epidemiol.* **36** 463–471. https://doi.org/10.1002/gepi.21640

Gong, Y. and Zou, F. (2012). Varying coefficient models for mapping quantitative trait loci using recombinant inbred intercrosses. *Genetics* **190** 475–486. https://doi.org/10.1534/genetics.111.132522

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. Ser. B* **55** 757–796. MR1229881

Hayeck, T. J., Zaitlen, N. A., Loh, P.-R., Vilhjalmsson, B., Pollack, S., Gusev, A. et al. (2015). Mixed model with correction for case-control ascertainment increases association power. *Am. J. Hum. Genet.* **96** 720–730.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822. MR1666699 https://doi.org/10.1093/biomet/85.4.809

Howie, B. N., Donnelly, P. and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5** e1000529. https://doi.org/10.1371/journal.pgen.1000529

Huang, J. Z. and Shen, H. (2004). Functional coefficient regression models for non-linear time series: A polynomial spline approach. *Scand. J. Stat.* **31** 515–534. MR2101537 https://doi.org/10.1111/j.1467-9469.2004.00404.x

Huang, J. Z., Wu, C. O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89** 111–128. MR1888349 https://doi.org/10.1093/biomet/89.1.111

Huang, J. Z., Wu, C. O. and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica* **14** 763–788. MR2087972

Jiang, D., Mbatchou, J. and McPeek, M. S. (2015). Retrospective association analysis of binary traits: Overcoming some limitations of the additive polygenic model. *Hum. Hered.* **80** 187–195. https://doi.org/10.1159/000446957

Jiang, D., Zhong, S. and McPeek, M. S. (2016). Retrospective binary-trait association test elucidates genetic architecture of Crohn disease. *Am. J. Hum. Genet.* **98** 243–255. https://doi.org/10.1016/j.ajhg.2015.12.012

Joe, H. (2015). *Dependence Modeling with Copulas*. *Monographs on Statistics and Applied Probability* **134**. CRC Press, Boca Raton, FL. MR3328438

Kauermann, G. and Tutz, G. (1999). On model diagnostics using varying coefficient models. *Biometrika* **86** 119–128. MR1688076 https://doi.org/10.1093/biomet/86.1.119

Kichaev, G., Bhatia, G., Loh, P. R., Gazal, S., Burch, K., Freund, M. K. et al. (2019). Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104** 65–75.

Kürüm, E., Hughes, J., Li, R. and Shiffman, S. (2018). Time-varying copula models for longitudinal data. *Stat. Interface* **11** 203–221. MR3780988 https://doi.org/10.4310/SII.2018.v11.n2.a1

KÜRÜM, E., LI, R., SHIFFMAN, S. and YAO, W. (2016). Time-varying coefficient models for joint modeling binary and continuous outcomes in longitudinal data. *Statist. Sinica* **26** 979–1000. MR3559939

LI, J., WANG, Z., LI, R. and WU, R. (2015). Bayesian group Lasso for nonparametric varying-coefficient models with application to functional genome-wide association studies. *Ann. Appl. Stat.* **9** 640–664. MR3371329 https://doi.org/10.1214/15-AOAS808

LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430 https://doi.org/10.1093/biomet/73.1.13

LIN, X. and BRESLOW, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *J. Amer. Statist. Assoc.* **91** 1007–1016. MR1424603 https://doi.org/10.2307/2291720

LIN, X. and CARROLL, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *J. Amer. Statist. Assoc.* **95** 520–534. MR1803170 https://doi.org/10.2307/2669396

LIN, X. and ZHANG, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 381–400. MR1680318 https://doi.org/10.1111/1467-9868.00183

LIU, H., TANG, Y. and ZHANG, H. H. (2009). A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Comput. Statist. Data Anal.* **53** 853–856. MR2657050 https://doi.org/10.1016/j.csda.2008.11.025

LIU, J., LI, R. and WU, R. (2014). Feature selection for varying coefficient models with ultrahigh-dimensional covariates. *J. Amer. Statist. Assoc.* **109** 266–274. MR3180562 https://doi.org/10.1080/01621459.2013.850086

LIU, Y. and XIE, J. (2020). Cauchy combination test: A powerful test with analytic p-value calculation under arbitrary dependency structures. *J. Amer. Statist. Assoc.* **115** 393–402. MR4078471 https://doi.org/10.1080/01621459.2018.1554485

LOIRAND, G. and PACAUD, P. (2010). The role of Rho protein signaling in hypertension. *Nat. Rev. Cardiol.* **7** 637–647. https://doi.org/10.1038/nrcardio.2010.136

LONDONO, D., CHEN, K., MUSOLF, A. et al. (2013). A novel method for analyzing genetic association with longitudinal phenotypes. *Stat. Appl. Genet. Mol. Biol.* **12** 241–261. MR3071030 https://doi.org/10.1515/sagmb-2012-0070

LU, Y. and ZHANG, R. (2009). Smoothing spline estimation of generalised varying-coefficient mixed model. *J. Nonparametr. Stat.* **21** 815–825. MR2572584 https://doi.org/10.1080/10485250903151078

MAACK, C., KARTES, T., KILTER, H., SCHÄFERS, H.-J., NICKENIG, G., BÖHM, M. et al. (2003). Oxygen free radical release in human failing myocardium is associated with increased activity of rac1-GTPase and represents a target for statin treatment. *Circulation* **108** 1567–1574.

MARCHETTI-BOWICK, M., YIN, J., HOWRYLAK, J. A. and XING, E. P. (2016). A time-varying group sparse additive model for genome-wide association studies of dynamic complex traits. *Bioinformatics* **32** 2903–2910.

MEIRELLES, O. D., DING, J., TANAKA, T., SANNA, S., YANG, H.-T., DUDEKULA, D. B. et al. (2013). SHAVE: Shrinkage estimator measured for multiple visits increases power in GWAS of quantitative traits. *Eur. J. Hum. Genet.* **21** 673–679.

MOUSTAFA-BAYOUMI, M., WISEL, S., GOLDSCHMIDT-CLERMONT, P. J. and HASSANAIN, H. H. (2003). Hypertension caused by transgenic overexpression of Rac1. *Med. Sci. Sports Exerc.* **35** S186.

NIKOLOULOPOULOS, A. K., JOE, H. and CHAGANTY, N. R. (2011). Eighted scores method for regression models with dependent data. *Biostatistics* **12** 653–665.

NING, C., KANG, H., ZHOU, L., WANG, D., WANG, H., WANG, A. et al. (2017). Performance gains in genome-wide association studies for longitudinal traits via modeling time-varied effects. *Sci. Rep.* **7**.

SATOH, M., OGITA, H., TAKESHITA, K., MUKAI, Y., KWIATKOWSKI, D. J. and LIAO, J. K. (2006). Requirement of Rac1 in the development of cardiac hypertrophy. *Proc. Natl. Acad. Sci. USA* **103** 7432–7437.

SAWADA, N., SALOMONE, S., KIM, H.-H., KWIATKOWSKI, D. J. and LIAO, J. K. (2008). Regulation of endothelial nitric oxide synthase and postnatal angiogenesis by Rac1. *Circ. Res.* **103** 360–368. https://doi.org/10.1161/CIRCRESAHA.108.178897

SCHAFFNER, S. F., FOO, C., GABRIEL, S., REICH, D., DALY, M. J. and ALTSHULER, D. (2005). Calibrating a coalescent simulation of human genome sequence variation. *Genome Res.* **15** 1576–1583. https://doi.org/10.1101/gr.3709305

SCHILDCROUT, J. S., SCHISTERMAN, E. F., MERCALDO, N. D., RATHOUZ, P. J. and HEAGERTY, P. J. (2018). Extending the case-control design to longitudinal data: Stratified sampling based on repeated binary outcomes. *Epidemiology* **29** 67–75. https://doi.org/10.1097/EDE.0000000000000764

SHAH, A., GRAY, K., FIGG, N., FINIGAN, A., STARKS, L. and BENNETT, M. (2018). Defective base excision repair of oxidative DNA damage in vascular smooth muscle cells promotes atherosclerosis. *Circulation* **138** 1446–1462. https://doi.org/10.1161/CIRCULATIONAHA.117.033249

SHAH, N. R. and MAHMOUDI, M. (2015). The role of DNA damage and repair in atherosclerosis: A review. *J. Mol. Cell. Cardiol.* **86** 147–157.

SHILOH, Y. and ZIV, Y. (2013). The ATM protein kinase: Regulating the cellular response to genotoxic stress, and more. *Nat. Rev., Mol. Cell Biol.* **14** 197–210.

SHLYAKHTER, I., SABETI, P. C. and SCHAFFNER, S. F. (2014). Cosi2: An efficient simulator of exact and approximate coalescent with selection. *Bioinformatics* **30** 3427–3429. https://doi.org/10.1093/bioinformatics/btu562

SIKORSKA, K., RIVADENEIRA, F., GROENEN, P. J. F., HOFMAN, A., UITTERLINDEN, A. G., EILERS, P. H. C. and LESAFFRE, E. (2013). Fast linear mixed model computations for genome-wide association studies with longitudinal data. *Stat. Med.* **32** 165–180. MR3017892 https://doi.org/10.1002/sim.5517

SITLANI, C. M., RICE, K. M., LUMLEY, T., MCKNIGHT, B., CUPPLES, L. A., AVERY, C. L., NOORDAM, R., STRICKER, B. H. C., WHITSEL, E. A. et al. (2015). Generalized estimating equations for genome-wide association studies using longitudinal phenotype data. *Stat. Med.* **34** 118–130. MR3286243 https://doi.org/10.1002/sim.6323

THORNTON, T. and MCPEEK, M. S. (2010). ROADTRIPS: Case-control association testing with partially or completely unknown population and pedigree structure. *Am. J. Hum. Genet.* **86** 172–184. https://doi.org/10.1016/j.ajhg.2010.01.001

URYGA, A., GRAY, K. and BENNETT, M. (2016). DNA damage and repair in vascular disease. *Annu. Rev. Physiol.* **78** 45–66. https://doi.org/10.1146/annurev-physiol-021115-105127

UZIEL, T., LERENTHAL, Y., MOYAL, L., ANDEGEKO, Y., MITTELMAN, L. and SHILOH, Y. (2003). Requirement of the MRN complex for ATM activation by DNA damage. *EMBO J.* **22** 5612–5621. https://doi.org/10.1093/emboj/cdg541

WAHBA, G. (1990). *Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442 https://doi.org/10.1137/1.9781611970128

WANG, L., LI, H. and HUANG, J. Z. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *J. Amer. Statist. Assoc.* **103** 1556–1569. MR2504204 https://doi.org/10.1198/016214508000000788

WANG, Z., XU, K., ZHANG, X., WU, X. and WANG, Z. (2017). Longitudinal SNP-set association analysis of quantitative phenotypes. *Genet. Epidemiol.* **41** 81–93.

WU, C. O., CHIANG, C.-T. and HOOVER, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **93** 1388–1402. MR1666635 https://doi.org/10.2307/2670054

WU, L., SOWERS, J. R., ZHANG, Y. and REN, J. (2022). Targeting DNA damage response in cardiovascular diseases: From pathophysiology to therapeutic implications. *Cardiovasc. Res.*

WU, W., WANG, Z., XU, K., ZHANG, X., AMEI, A., GELERNTER, J. et al. (2019). Retrospective association analysis of longitudinal binary traits identifies important loci and pathways in cocaine use. *Genetics* **213** 1225–1236.

WU, X. and MCPEEK, M. S. (2018). L-GATOR: Genetic association testing for a longitudinally measured quantitative trait in samples with related individuals. *Am. J. Hum. Genet.* **102** 574–591. https://doi.org/10.1016/j.ajhg.2018.02.016

XIA, X., YANG, H. and LI, J. (2016). Feature screening for generalized varying coefficient models with application to dichotomous responses. *Comput. Statist. Data Anal.* **102** 85–97. MR3506984 https://doi.org/10.1016/j.csda.2016.04.008

XU, G., AMEI, A., WU, W., LIU, Y., SHEN, L., OH, E. C and WANG, Z. (2024). Supplement to "Retrospective varying coefficient association analysis of longitudinal binary traits: Application to the identification of genetic loci associated with hypertension." https://doi.org/10.1214/23-AOAS1798SUPPA, https://doi.org/10.1214/23-AOAS1798SUPPB

ZHANG, D. (2004). Generalized linear mixed models with varying coefficients for longitudinal data. *Biometrics* **60** 8–15. MR2043613 https://doi.org/10.1111/j.0006-341X.2004.00165.x

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 506–528
https://doi.org/10.1214/23-AOAS1799

WHAT MAKES FOREST-BASED HETEROGENEOUS TREATMENT EFFECT ESTIMATORS WORK?

BY SUSANNE DANDL[1,a], CHRISTIAN HASLINGER[2,b], TORSTEN HOTHORN[3,c], HEIDI SEIBOLD[4,d], ERIK SVERDRUP[5,e], STEFAN WAGER[5,f] AND ACHIM ZEILEIS[6,g]

[1] *Institut für Statistik, Ludwig-Maximilians-Universität München,* [a] *Susanne.Dandl@stat.uni-muenchen.de*
[2] *Klinik für Geburtshilfe, Universitätsspital und Universität Zürich,* [b] *Christian.Haslinger@usz.ch*
[3] *Institut für Epidemiologie, Biostatistik und Prävention, Universität Zürich,* [c] *Torsten.Hothorn@R-project.org*
[4] *Institute for Globally Distributed Open Research and Education (IGDORE), München,* [d] *heidi@seibold.co*
[5] *Stanford Graduate School of Business, Stanford University,* [e] *erikcs@stanford.edu,* [f] *swager@stanford.edu*
[6] *Faculty of Economics and Statistics, Universität Innsbruck,* [g] *Achim.Zeileis@R-project.org*

Estimation of heterogeneous treatment effects (HTE) is of prime importance in many disciplines, from personalized medicine to economics among many others. Random forests have been shown to be a flexible and powerful approach to HTE estimation in both randomized trials and observational studies. In particular "causal forests" introduced by Athey, Tibshirani and Wager (*Ann. Statist.* **47** (2019) 1148–1178), along with the R implementation in package *grf* were rapidly adopted. A related approach, called "model-based forests" that is geared toward randomized trials and simultaneously captures effects of both prognostic and predictive variables, was introduced by Seibold, Zeileis and Hothorn (*Stat. Methods Med. Res.* **27** (2018) 3104–3125) along with a modular implementation in the R package *model4you*.

Neither procedure is directly applicable to the estimation of individualized predictions of excess postpartum blood loss caused by a cesarean section in comparison to vaginal delivery. Clearly, randomization is hardly possible in this setup, and thus model-based forests lack clinical trial data to address this question. On the other hand, the skewed and interval-censored postpartum blood loss observations violate assumptions made by causal forests. Here we present a tailored model-based forest for skewed and interval-censored data to infer possible predictive prepartum characteristics and their impact on excess postpartum blood loss caused by a cesarean section.

As a methodological basis, we propose a unifying view on causal and model-based forests that goes beyond the *theoretical* motivations and investigates which *computational* elements make causal forests so successful and how these can be blended with the strengths of model-based forests. To do so, we show that both methods can be understood in terms of the same parameters and model assumptions for an additive model under L_2 loss. This theoretical insight allows us to implement several flavors of "model-based causal forests" and dissect their different elements *in silico*.

The original causal forests and model-based forests are compared with the new blended versions in a benchmark study exploring both randomized trials and observational settings. In the randomized setting, both approaches performed akin. If confounding was present in the data-generating process, we found local centering of the treatment indicator with the corresponding propensities to be the main driver for good performance. Local centering of the outcome was less important and might be replaced or enhanced by simultaneous split selection with respect to both prognostic and predictive effects. This lays the foundation for future research combining random forests for HTE estimation with other types of models.

Key words and phrases. Causal forests, heterogeneous treatment effects, observational data, personalized medicine, postpartum hemorrhage, random forest.

REFERENCES

AKAZAWA, M., HASHIMOTO, K., KATSUHIKO, N. and KANAME, Y. (2021). Machine learning approach for the prediction of postpartum hemorrhage in vaginal birth. *Sci. Rep.* **11** 22620. https://doi.org/10.1038/s41598-021-02198-y

ANTOINE, C. and YOUNG, B. K. (2021). Cesarean section one hundred years 1920–2020: The good, the bad and the ugly. *J. Perinat. Med.* **49** 5–16. https://doi.org/10.1515/jpm-2020-0305

ATHEY, S. and IMBENS, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. USA* **113** 7353–7360. MR3531135 https://doi.org/10.1073/pnas.1510489113

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

ATHEY, S. and WAGER, S. (2019). Estimating treatment effects with causal forests: An application. *Obs. Stud.* **5** 37–51. https://doi.org/10.1353/obs.2019.0001

BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32. https://doi.org/10.1023/a:1010933404324

BROOKS, M., BOLKER, B., KRISTENSEN, K., MAECHLER, M., MAGNUSSON, A., SKAUG, H., NIELSEN, A., BERG, C. and VAN BENTHAM, K. (2021). *glmmTMB*: Generalized linear mixed models using template model builder. R package version 1.1.2. Available at https://CRAN.R-project.org/package=glmmTMB.

BURI, M. and HOTHORN, T. (2020). Model-based random forests for ordinal regression. *Int. J. Biostat.* **16** 20190063. https://doi.org/10.1515/ijb-2019-0063

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

CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4** 266–298. MR2758172 https://doi.org/10.1214/09-AOAS285

CUI, Y., KOSOROK, M. R., SVERDRUP, E., WAGER, S. and ZHU, R. (2023). Estimating heterogeneous treatment effects with right-censored data via causal survival forests. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 179–211. https://doi.org/10.1093/jrsssb/qkac001

DANDL, S., HASLINGER, C., HOTHORN, T., SEIBOLD, H., SVERDRUP, E., WAGER, S. and ZEILEIS, A. (2024a). Details and additional results. Supplement A to "What makes forest-based heterogeneous treatment effect estimators work?" https://doi.org/10.1214/23-AOAS1799SUPPA

DANDL, S., HASLINGER, C., HOTHORN, T., SEIBOLD, H., SVERDRUP, E., WAGER, S. and ZEILEIS, A. (2024b). Code and data. Supplement B to "What makes forest-based heterogeneous treatment effect estimators work?" https://doi.org/10.1214/23-AOAS1799SUPPB

DASGUPTA, A., SZYMCZAK, S., MOORE, J. H., BAILEY-WILSON, J. E. and MALLEY, J. D. (2014). Risk estimation using probability machines. *BioData Min.* **7** 2. https://doi.org/10.1186/1756-0381-7-2

ENDE, H. B. (2022). Risk assessment tools to predict postpartum hemorrhage. *Bailliere's Best Pract. Res., Clin. Anaesthesiol.* **36** 341–348. https://doi.org/10.1016/j.bpa.2022.08.003

ERICKSON, E. N. and CARLSON, N. S. (2020). Predicting postpartum hemorrhage after low-risk vaginal birth by labor characteristics and oxytocin administration. *J. Obstet. Gynecol. Neonatal Nurs.* **49** 549–563. https://doi.org/10.1016/j.jogn.2020.08.005

FAROUKI, R. T. (2012). The Bernstein polynomial basis: A centennial retrospective. *Comput. Aided Geom. Design* **29** 379–419. MR2921860 https://doi.org/10.1016/j.cagd.2012.03.001

FENG, P., ZHOU, X.-H., ZOU, Q.-M., FAN, M.-Y. and LI, X.-S. (2012). Generalized propensity score for estimating the average treatment effect of multiple treatments. *Stat. Med.* **31** 681–697. MR2900870 https://doi.org/10.1002/sim.4168

FOKKEMA, M., SMITS, N., ZEILEIS, A., HOTHORN, T. and KELDERMAN, H. (2018). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behav. Res. Methods* **50** 2016–2034. https://doi.org/10.3758/s13428-017-0971-x

FOSTER, J. C., TAYLOR, J. M. G. and RUBERG, S. J. (2011). Subgroup identification from randomized clinical trial data. *Stat. Med.* **30** 2867–2880. MR2844689 https://doi.org/10.1002/sim.4322

HASLINGER, C., KORTE, W., HOTHORN, T., BRUN, R., GREENBERG, C. and ZIMMERMANN, R. (2020). The impact of prepartum factor XIII activity on postpartum blood loss. *J. Thromb. Haemost.* **18** 1310–1319. https://doi.org/10.1111/jth.14795

HILL, J. L. (2011). Bayesian nonparametric modeling for causal inference. *J. Comput. Graph. Statist.* **20** 217–240. MR2816546 https://doi.org/10.1198/jcgs.2010.08162

HOTHORN, T. (2021). *trtf*: Transformation trees and forests. R package version 0.3-8. Available at https://CRAN.R-project.org/package=trtf.

HOTHORN, T., BRETZ, F. and WESTFALL, P. (2021). *multcomp*: Simultaneous inference in general parametric models. R package version 1.4-17. Available at https://CRAN.R-project.org/package=multcomp.

HOTHORN, T., HORNIK, K. and ZEILEIS, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *J. Comput. Graph. Statist.* **15** 651–674. MR2291267 https://doi.org/10.1198/106186006X133933

HOTHORN, T., LAUSEN, B., BENNER, A. and RADESPIEL-TRÖGER, M. (2004). Bagging survival trees. *Stat. Med.* **23** 77–91. https://doi.org/10.1002/sim.1593

HOTHORN, T., MÖST, L. and BÜHLMANN, P. (2018). Most likely transformations. *Scand. J. Stat.* **45** 110–134. MR3764288 https://doi.org/10.1111/sjos.12291

HOTHORN, T. and ZEILEIS, A. (2015). *partykit*: A modular toolkit for recursive partytioning in R. *J. Mach. Learn. Res.* **16** 3905–3909. MR3450556

HOTHORN, T. and ZEILEIS, A. (2021a). Predictive distribution modeling using transformation forests. *J. Comput. Graph. Statist.* **30** 1181–1196. MR4356613 https://doi.org/10.1080/10618600.2021.1872581

HOTHORN, T. and ZEILEIS, A. (2021b). *partykit*: A toolkit for recursive partytioning. R package version 1.2-15. Available at https://CRAN.R-project.org/package=partykit.

IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—For Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. MR3309951 https://doi.org/10.1017/CBO9781139025751

ISHWARAN, H., KOGALUR, U. B., BLACKSTONE, E. H. and LAUER, M. S. (2008). Random survival forests. *Ann. Appl. Stat.* **2** 841–860. MR2516796 https://doi.org/10.1214/08-AOAS169

ISHWARAN, H. and MALLEY, J. D. (2014). Synthetic learning machines. *BioData Min.* **7** 28. https://doi.org/10.1186/s13040-014-0028-y

KAHR, M. K., BRUN, R., ZIMMERMANN, R., FRANKE, D. and HASLINGER, C. (2018). Validation of a quantitative system for real-time measurement of postpartum blood loss. *Arch. Gynecol. Obstet.* **298** 1071–1077. https://doi.org/10.1007/s00404-018-4896-0

KAWAKITA, T., MOKHTARI, N., HUANG, J. C. and LANDY, H. J. (2019). Evaluation of risk-assessment tools for severe postpartum hemorrhage in women undergoing Cesarean delivery. *Obstet. Gynecol.* **134** 1308–1316. https://doi.org/10.1097/AOG.0000000000003574

KOREPANOVA, N., SEIBOLD, H., STEFFEN, V. and HOTHORN, T. (2020). Survival forests under test: Impact of the proportional hazards assumption on prognostic and predictive forests for amyotrophic lateral sclerosis survival. *Stat. Methods Med. Res.* **29** 1403–1419. MR4097153 https://doi.org/10.1177/0962280219862586

KÜNZEL, S. R., SEKHON, J. S., BICKEL, P. J. and YU, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Natl. Acad. Sci. USA* **116** 4156–4165. https://doi.org/10.1073/pnas.1804597116

LIN, Y. and JEON, Y. (2006). Random forests and adaptive nearest neighbors. *J. Amer. Statist. Assoc.* **101** 578–590. MR2256176 https://doi.org/10.1198/016214505000001230

LU, M., SADIQ, S., FEASTER, D. J. and ISHWARAN, H. (2018). Estimating individual treatment effect in observational data using random forest methods. *J. Comput. Graph. Statist.* **27** 209–219. MR3788313 https://doi.org/10.1080/10618600.2017.1356325

MACDORMAN, M. F., DECLERCQ, E., CABRAL, H. and MORTON, C. (2016). Recent increases in the U.S. maternal mortality rate: Disentangling trends from measurement issues. *Obstet. Gynecol.* **128** 447–455. https://doi.org/10.1097/AOG.0000000000001556

MAYER, I., SVERDRUP, E., GAUSS, T., MOYER, J.-D., WAGER, S. and JOSSE, J. (2020). Doubly robust treatment effect estimation with missing attributes. *Ann. Appl. Stat.* **14** 1409–1431. MR4152139 https://doi.org/10.1214/20-AOAS1356

MEINSHAUSEN, N. (2006). Quantile regression forests. *J. Mach. Learn. Res.* **7** 983–999. MR2274394

NIE, X. and WAGER, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* **108** 299–319. MR4259133 https://doi.org/10.1093/biomet/asaa076

POWERS, S., QIAN, J., JUNG, K., SCHULER, A., SHAH, N. H., HASTIE, T. and TIBSHIRANI, R. (2018). Some methods for heterogeneous treatment effect estimation in high dimensions. *Stat. Med.* **37** 1767–1787. MR3799840 https://doi.org/10.1002/sim.7623

PROBST, P., BOULESTEIX, A.-L. and BISCHL, B. (2019). Tunability: Importance of hyperparameters of machine learning algorithms. *J. Mach. Learn. Res.* **20** Paper No. 53, 32 pp. MR3948093

PROBST, P., WRIGHT, M. N. and BOULESTEIX, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Min. Knowl. Discov.* **9** e1301. https://doi.org/10.1002/widm.1301

R CORE TEAM (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org.

ROBINSON, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56** 931–954. MR0951762 https://doi.org/10.2307/1912705

ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

SAY, L., CHOU, D., GEMMILL, A., TUNÇALP, O., MOLLER, A.-B., DANIELS, J., GÜLMEZOGLU, A. M., TEMMERMAN, M. and ALKEMA, L. (2014). Global causes of maternal death: A WHO systematic analysis. *Lancet Glob. Health* **2** e323–e333. https://doi.org/10.1016/S2214-109X(14)70227-X

SCHLOSSER, L., HOTHORN, T., STAUFFER, R. and ZEILEIS, A. (2019). Distributional regression forests for probabilistic precipitation forecasting in complex terrain. *Ann. Appl. Stat.* **13** 1564–1589. MR4019150 https://doi.org/10.1214/19-AOAS1247

SCHNELL, P., TANG, Q., MÜLLER, P. and CARLIN, B. P. (2017). Subgroup inference for multiple treatments and multiple endpoints in an Alzheimer's disease treatment trial. *Ann. Appl. Stat.* **11** 949–966. MR3693553 https://doi.org/10.1214/17-AOAS1024

SEIBOLD, H., ZEILEIS, A. and HOTHORN, T. (2016). Model-based recursive partitioning for subgroup analyses. *Int. J. Biostat.* **12** 45–63. MR3505686 https://doi.org/10.1515/ijb-2015-0032

SEIBOLD, H., ZEILEIS, A. and HOTHORN, T. (2018). Individual treatment effect prediction for amyotrophic lateral sclerosis patients. *Stat. Methods Med. Res.* **27** 3104–3125. MR3855657 https://doi.org/10.1177/0962280217693034

SEIBOLD, H., ZEILEIS, A. and HOTHORN, T. (2019). *model4you*: An R package for personalised treatment effect estimation. *J. Open Res. Softw.* **7** 1–6. https://doi.org/10.5334/jors.219

SEIBOLD, H., ZEILEIS, A. and HOTHORN, T. (2021). *model4you*: Stratified and personalised models based on model-based trees and forests. R package version 0.9-7. Available at https://CRAN.R-project.org/package=model4you.

STARLING, J. E., MURRAY, J. S., LOHR, P. A., AIKEN, A. R. A., CARVALHO, C. M. and SCOTT, J. G. (2021). Targeted smooth Bayesian causal forests: An analysis of heterogeneous treatment effects for simultaneous vs. interval medical abortion regimens over gestation. *Ann. Appl. Stat.* **15** 1194–1219. MR4316646 https://doi.org/10.1214/20-aoas1438

TANG, F. and ISHWARAN, H. (2017). Random forest missing data algorithms. *Stat. Anal. Data Min.* **10** 363–377. MR3733611 https://doi.org/10.1002/sam.11348

TIBSHIRANI, J., ATHEY, S., SVERDRUP, E. and WAGER, S. (2021). *grf*: Generalized random forests. R package version 2.0.2. Available at https://CRAN.R-project.org/package=grf.

VENKATESH, K. K., STRAUSS, R. A., GROTEGUT, C. A., HEINE, R. P., CHESCHEIR, N. C., STRINGER, J. S. A., STAMILIO, D. M., MENARD, K. M. and JELOVSEK, J. E. (2020). Machine learning and statistical models to predict postpartum hemorrhage. *Obstet. Gynecol.* **135** 935–944. https://doi.org/10.1097/AOG.0000000000003759

WAGER, S. and ATHEY, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* **113** 1228–1242. MR3862353 https://doi.org/10.1080/01621459.2017.1319839

WHO (2012). WHO recommendations for the prevention and treatment of postpartum haemorrhage. World Health Organization, Geneva, Switzerland.

ZANUTTO, E., LU, B. and HORNIK, R. (2005). Using propensity score subclassification for multiple treatment doses to evaluate a national antidrug media campaign. *J. Educ. Behav. Stat.* **30** 59–73. https://doi.org/10.3102/10769986030001059

ZEILEIS, A., HOTHORN, T. and HORNIK, K. (2008). Model-based recursive partitioning. *J. Comput. Graph. Statist.* **17** 492–514. MR2439970 https://doi.org/10.1198/106186008X319331

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 529–554
https://doi.org/10.1214/23-AOAS1800

A NONSEPARABLE FIRST-ORDER SPATIOTEMPORAL INTENSITY FOR EVENTS ON LINEAR NETWORKS: AN APPLICATION TO AMBULANCE INTERVENTIONS

BY ANDREA GILARDI[1,a], RICCARDO BORGONI[1,b] AND JORGE MATEU[2,c]

[1]*Department of Economics, Management and Statistics, University of Milan—Bicocca,* [a]*andrea.gilardi@unimib.it,* [b]*riccardo.borgoni@unimib.it*

[2]*Department of Mathematics, Universitat Jaume I,* [c]*mateu@uji.es*

The algorithms used for the optimal management of an ambulance fleet require an accurate description of the spatiotemporal evolution of the emergency events. In the last years, several authors have proposed sophisticated statistical approaches to forecast ambulance dispatches, typically modelling the data as a point pattern occurring on a planar region. Nevertheless, ambulance interventions can be more appropriately modelled as a realisation of a point process occurring on a linear network. The constrained spatial domain raises specific challenges and unique methodological problems that cannot be ignored when developing a proper statistical approach. Hence, this paper proposes a spatiotemporal model to analyse ambulance dispatches focusing on the interventions that occurred in the road network of Milan (Italy) from 2015 to 2017. We adopt a nonseparable first-order intensity function with spatial and temporal terms. The temporal dimension is estimated semiparametrically using a Poisson regression model, while the spatial dimension is estimated nonparametrically using a network kernel function. A set of weights is included in the spatial term to capture space-time interactions, inducing nonseparability in the intensity function. A series of tests show that our approach successfully models the ambulance interventions and captures the space-time patterns more accurately than planar or separable point process models.

REFERENCES

ANG, Q. W., BADDELEY, A. and NAIR, G. (2012). Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scand. J. Stat.* **39** 591–617. MR3000837 https://doi.org/10.1111/j.1467-9469.2011.00752.x

BADDELEY, A., RUBAK, E. and TURNER, R. (2015). *Spatial Point Patterns*: *Methodology and Applications with R*. CRC Press, Boca Raton, FL.

BADDELEY, A., NAIR, G., RAKSHIT, S., MCSWIGGAN, G. and DAVIES, T. M. (2021). Analysing point patterns on networks—A review. *Spat. Stat.* **42** Paper No. 100435, 35 pp. MR4233256 https://doi.org/10.1016/j.spasta.2020.100435

BARRINGTON-LEIGH, C. and MILLARD-BALL, A. (2017). The world's user-generated road map is more than 80% complete. *PLoS ONE* **12** e0180698. https://doi.org/10.1371/journal.pone.0180698

BARTHÉLEMY, M. (2011). Spatial networks. *Phys. Rep.* **499** 1–101. MR2770962 https://doi.org/10.1016/j.physrep.2010.11.002

BAYISA, F. L., ÅDAHL, M., RYDÉN, P. and CRONIE, O. (2020). Large-scale modelling and forecasting of ambulance calls in northern Sweden using spatio-temporal log-Gaussian Cox processes. *Spat. Stat.* **39** Paper No. 100471, 22 pp. MR4158686 https://doi.org/10.1016/j.spasta.2020.100471

BLACKWELL, T. H. and KAUFMAN, J. S. (2002). Response time effectiveness: Comparison of response time and survival in an urban emergency medical services system. *Acad. Emerg. Med.* **9** 288–295.

BYRD, R. H., LU, P., NOCEDAL, J. and ZHU, C. Y. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **16** 1190–1208. MR1346301 https://doi.org/10.1137/0916069

OPENSTREETMAP CONTRIBUTORS (2017). Planet dump. Available at https://planet.osm.org. https://www.openstreetmap.org.

Key words and phrases. Emergency interventions, linear networks, nonparametric methods, point patterns on linear networks, spatial networks, spatiotemporal data.

DIGGLE, P., ROWLINGSON, B. and SU, T. (2005). Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics* **16** 423–434. MR2147534 https://doi.org/10.1002/env.712

DIGGLE, P. J. (2014). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, 3rd ed. *Monographs on Statistics and Applied Probability* **128**. CRC Press, Boca Raton, FL. MR3113855

DIGGLE, P. J. (1985). A kernel method for smoothing point process data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **34** 138–147.

KELSALL, J. E. and DIGGLE, P. J. (1998). Spatial variation in risk of disease: A nonparametric binary regression approach. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **47** 559–573.

GILARDI, A., BORGONI, R. and MATEU, J. (2024). Supplement to "A nonseparable first-order spatiotemporal intensity for events on linear networks: An application to ambulance interventions." https://doi.org/10.1214/23-AOAS1800SUPP

GILARDI, A. and LOVELACE, R. (2021). osmextract: Download and read OpenStreetMap data extracts.

GONZÁLEZ, J. A., RODRÍGUEZ-CORTÉS, F. J., CRONIE, O. and MATEU, J. (2016). Spatio-temporal point process statistics: A review. *Spat. Stat.* **18** 505–544. MR3575505 https://doi.org/10.1016/j.spasta.2016.10.002

HENDERSON, S. G. (2011). Operations research tools for addressing current challenges in emergency medical services. In *Wiley Encyclopedia of Operations Research and Management Science* Wiley, Inc, Hoboken, NJ. https://doi.org/10.1002/9780470400531.eorms0605

JALILIAN, A. and MATEU, J. (2023). Assessing similarities between spatial point patterns with a Siamese neural network discriminant model. *Adv. Data Anal. Classif.* **17** 21–42. MR4552824 https://doi.org/10.1007/s11634-021-00485-0

JONES, M. C. (1993). Simple boundary correction for kernel density estimation. *Stat. Comput.* **3** 135–146.

KELSALL, J. E. and DIGGLE, P. J. (1995). Kernel estimation of relative risk. *Bernoulli* **1** 3–16. MR1354453 https://doi.org/10.2307/3318678

LU, Y. and CHEN, X. (2007). On the false alarm of planar K-function when analyzing urban crime distributed along streets. *Soc. Sci. Res.* **36** 611–632.

MATEU, J. and JALILIAN, A. (2022). Spatial point processes and neural networks: A convenient couple. *Spat. Stat.* **50** Paper No. 100644, 19 pp. MR4439344 https://doi.org/10.1016/j.spasta.2022.100644

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 https://doi.org/10.1214/10-AOAS442

MCSWIGGAN, G., BADDELEY, A. and NAIR, G. (2017). Kernel density estimation on a linear network. *Scand. J. Stat.* **44** 324–345. MR3658517 https://doi.org/10.1111/sjos.12255

MCSWIGGAN, G., BADDELEY, A. and NAIR, G. (2020). Estimation of relative risk for events on a linear network. *Stat. Comput.* **30** 469–484. MR4064631 https://doi.org/10.1007/s11222-019-09889-7

MØLLER, J., SYVERSVEEN, A. R. and WAAGEPETERSEN, R. P. (1998). Log Gaussian Cox processes. *Scand. J. Stat.* **25** 451–482. MR1650019 https://doi.org/10.1111/1467-9469.00115

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

OKABE, A. and SUGIHARA, K. (2012). *Spatial Analysis Along Networks: Statistical and Computational Methods*. Wiley, New York.

RAKSHIT, S., BADDELEY, A. and NAIR, G. (2019). Efficient code for second order analysis of events on a linear network. *J. Stat. Softw.* **90** 1–37.

RAKSHIT, S., DAVIES, T., MORADI, M. M., MCSWIGGAN, G., NAIR, G., MATEU, J. and BADDELEY, A. (2019). Fast kernel smoothing of point patterns on a large network using two-dimensional convolution. *Int. Stat. Rev.* **87** 531–556. MR4043351 https://doi.org/10.1111/insr.12327

SCOTT, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR1191168 https://doi.org/10.1002/9780470316849

SILVERMAN, B. W. (1982). Algorithm AS 176: Kernel density estimation using the fast Fourier transform. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **31** 93–99.

R CORE TEAM (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

VILE, J. L., GILLARD, J. W., HARPER, P. R. and KNIGHT, V. A. (2012). Predicting ambulance demand using singular spectrum analysis. *J. Oper. Res. Soc.* **63** 1556–1565.

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 https://doi.org/10.1111/j.1467-9868.2010.00749.x

WOOD, S. N. (2017). *Generalized Additive Models: An Introduction with R*, 2 ed. Chapman and Hall/CRC, Boca Raton, FL.

YAMADA, I. and THILL, J.-C. (2004). Comparison of planar and network K-functions in traffic accident analysis. *J. Transp. Geogr.* **12** 149–158.

YUAN, B., WANG, X., MA, J., ZHOU, C., BERTOZZI, A. L. and YANG, H. (2019). Variational autoencoders for highly multivariate spatial point processes intensities. In *International Conference on Learning Representations*.

ZHOU, Z. and MATTESON, D. S. (2015). Predicting ambulance demand: A spatio-temporal kernel approach. In *Proceedings of the* 21*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. *KDD* '15 2297–2303. Association for Computing Machinery, New York, NY, USA.

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 https://doi.org/10.1080/01621459.2014.941466

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 555–584
https://doi.org/10.1214/23-AOAS1801

CHANGE POINT DETECTION IN DYNAMIC GAUSSIAN GRAPHICAL MODELS: THE IMPACT OF COVID-19 PANDEMIC ON THE U.S. STOCK MARKET

BY BEATRICE FRANZOLINI[1,a], ALEXANDROS BESKOS[2,b], MARIA DE IORIO[3,d], WARRICK POKLEWSKI KOZIELL[2,c] AND KAROLINA GRZESZKIEWICZ[4,e]

[1]*Bocconi Institute for Data Science and Analytics, Bocconi University,* [a]*beatrice.franzolini@unibocconi.it*
[2]*Department of Statistical Science, University College London,* [b]*a.beskos@ucl.ac.uk,* [c]*warrickpk@gmail.com*
[3]*Yong Loo Lin School of Medicine, National University of Singapore,* [d]*mdi@nus.edu.sg*
[4]*Yale-NUS College,* [e]*karolina.grzeszkiewicz@u.yale-nus.edu.sg*

Reliable estimates of volatility and correlation are fundamental in economics and finance for understanding the impact of macroeconomics events on the market and guiding future investments and policies. Dependence across financial returns is likely to be subject to sudden structural changes, especially in correspondence with major global events, such as the COVID-19 pandemic. In this work we are interested in capturing abrupt changes over time in the conditional dependence across U.S. industry stock portfolios, over a time horizon that covers the COVID-19 pandemic. The selected stocks give a comprehensive picture of the U.S. stock market. To this end, we develop a Bayesian multivariate stochastic volatility model based on a time-varying sequence of graphs capturing the evolution of the dependence structure. The model builds on the Gaussian graphical models and the random change points literature. In particular, we treat the number, the position of change points, and the graphs as object of posterior inference, allowing for sparsity in graph recovery and change point detection. The high dimension of the parameter space poses complex computational challenges. However, the model admits a hidden Markov model formulation. This leads to the development of an efficient computational strategy, based on a combination of sequential Monte-Carlo and Markov chain Monte-Carlo techniques. Model and computational development are widely applicable, beyond the scope of the application of interest in this work.

REFERENCES

AB RAZAK, R., AMINUDDIN, N. H. and ISMAIL, N. (2018). Dependence measure of daily versus weekly returns. *Int. J. Eng. Technol.* **7** 329–333.

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

ALQARALLEH, H. and CANEPA, A. (2021). Evidence of stock market contagion during the COVID-19 pandemic: A Wavelet-Copula-GARCH approach. *J. Risk Financ. Manag.* **14** 329.

ANASTASIOU, A., CRIBBEN, I. and FRYZLEWICZ, P. (2022). Cross-covariance isolate detect: A new change-point method for estimating dynamic functional connectivity. *Med. Image Anal.* **75** 102252.

ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. MR2758115 https://doi.org/10.1111/j.1467-9868.2009.00736.x

ASAI, M., MCALEER, M. and YU, J. (2006). Multivariate stochastic volatility: A review. *Econometric Rev.* **25** 145–175. MR2256285 https://doi.org/10.1080/07474930600713564

ATAY-KAYIS, A. and MASSAM, H. (2005). A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika* **92** 317–335. MR2201362 https://doi.org/10.1093/biomet/92.2.317

BANERJEE, S. and GHOSAL, S. (2015). Bayesian structure learning in graphical models. *J. Multivariate Anal.* **136** 147–162. MR3321485 https://doi.org/10.1016/j.jmva.2015.01.015

Key words and phrases. Coronavirus pandemic, graphical models, industry portoflios, particle filter, precision matrix, stochastic volatility.

BATES, D., MAECHLER, M., JAGAN, M. and DAVIS, T. A. (2022). Matrix: Sparse and dense matrix classes and methods, R package version 1.5-3.

BAUWENS, L., LAURENT, S. and ROMBOUTS, J. V. K. (2006). Multivariate GARCH models: A survey. *J. Appl. Econometrics* **21** 79–109. MR2225523 https://doi.org/10.1002/jae.842

BESKOS, A., CRISAN, D. and JASRA, A. (2014). On the stability of sequential Monte Carlo methods in high dimensions. *Ann. Appl. Probab.* **24** 1396–1445. MR3211000 https://doi.org/10.1214/13-AAP951

BIANCHI, D., BILLIO, M., CASARIN, R. and GUIDOLIN, M. (2019). Modeling systemic risk with Markov switching graphical SUR models. *J. Econometrics* **210** 58–74. MR3944763 https://doi.org/10.1016/j.jeconom.2018.11.005

BILMES, J. A. (2004). Graphical models and automatic speech recognition. In *Mathematical Foundations of Speech and Language Processing*. *IMA Vol. Math. Appl.* **138** 191–245. Springer, New York. MR2074542 https://doi.org/10.1007/978-1-4419-9017-4_10

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327. MR0853051 https://doi.org/10.1016/0304-4076(86)90063-1

BOLLERSLEV, T., ENGLE, R. F. and NELSON, D. B. (1994). Arch models. In *Handbook of Econometrics, Vol. IV*. *Handbooks in Econom.* **2** 2959–3038. North-Holland, Amsterdam. MR1315984

BOUDT, K., GALANOS, A., PAYSEUR, S. and ZIVOT, E. (2019). Multivariate GARCH models for large-scale applications: A survey. In *Handbook of Statistics* **41** 193–242. Elsevier.

BOUVIER, A., GIRAUD, C., HUET, S. and VERZELEN, N. (2022). GGMselect: Gaussian graphs models selection. R package version 0.1-12.5.

BYBEE, L. and ATCHADÉ, Y. (2018). Change-point computation for large graphical models: A scalable algorithm for Gaussian graphical models with change-points. *J. Mach. Learn. Res.* **19** Paper No. 11, 38. MR3862418

CAPORALE, G. M. and ZEKOKH, T. (2019). Modelling volatility of cryptocurrencies using Markov-switching GARCH models. *Res. Int. Bus. Finance* **48** 143–155.

CARVALHO, C. M., MASSAM, H. and WEST, M. (2007). Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika* **94** 647–659. MR2410014 https://doi.org/10.1093/biomet/asm056

CARVALHO, C. M. and WEST, M. (2007). Dynamic matrix-variate graphical models. *Bayesian Anal.* **2** 69–97. MR2289924 https://doi.org/10.1214/07-BA204

CHAN, J. C. C. (2013). Moving average stochastic volatility models with application to inflation forecast. *J. Econometrics* **176** 162–172. MR3084051 https://doi.org/10.1016/j.jeconom.2013.05.003

CHANDRA, N. K., MUELLER, P. and SARKAR, A. (2021). Bayesian scalable precision factor analysis for massive sparse Gaussian graphical models. arXiv preprint, arXiv:2107.11316.

CHENG, Y. and LENKOSKI, A. (2012). Hierarchical Gaussian graphical models: Beyond reversible jump. *Electron. J. Stat.* **6** 2309–2331. MR3020264 https://doi.org/10.1214/12-EJS746

CHIB, S. and GREENBERG, E. (1998). Analysis of multivariate probit models. *Biometrika* **85** 347–361.

CHOPIN, N. and PAPASPILIOPOULOS, O. (2020). *An Introduction to Sequential Monte Carlo*. *Springer Series in Statistics*. Springer, Cham. MR4215639 https://doi.org/10.1007/978-3-030-47845-2

CLARK, T. E. and RAVAZZOLO, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *J. Appl. Econometrics* **30** 551–575. MR3358636 https://doi.org/10.1002/jae.2379

CONTESSI, S., DE PACE, P. and GUIDOLIN, M. (2014). How did the financial crisis alter the correlations of US yield spreads? *J. Empir. Finance* **28** 362–385.

CRIBBEN, I. (2019). Change points in heavy-tailed multivariate time series: Methods using precision matrices. *Appl. Stoch. Models Bus. Ind.* **35** 299–320. MR3946128 https://doi.org/10.1002/asmb.2373

CRIBBEN, I. and YU, Y. (2017). Estimating whole-brain dynamics by using spectral clustering. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 607–627. MR3632344 https://doi.org/10.1111/rssc.12169

CUI, K. (2012). Semiparametric Gaussian variance-mean mixtures for heavy-tailed and skewed data. *Int. Sch. Res. Not.* **2012**.

DANAHER, P., WANG, P. and WITTEN, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 373–397. MR3164871 https://doi.org/10.1111/rssb.12033

DAWID, A. P. and LAURITZEN, S. L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272–1317. MR1241267 https://doi.org/10.1214/aos/1176349260

DEL MORAL, P. (2004). *Feynman–Kac Formulae*: *Genealogical and Interacting Particle Systems with Applications*. *Probability and Its Applications* (*New York*). Springer, New York. MR2044973 https://doi.org/10.1007/978-1-4684-9393-1

DELLAPORTAS, P., GIUDICI, P. and ROBERTS, G. (2003). Bayesian inference for nondecomposable graphical Gaussian models. *Sankhyā* **65** 43–55. MR2016776

DERBALI, A., NAOUI, K., SASSI, M. B. and AMIRI, M. M. (2022). Do COVID-19 epidemic explains the dynamic conditional correlation between China's stock market index and international stock market indices? *Chin. Econ.* **55** 227–242.

DEY, A. K., HOQUE, G. T., DAS, K. P. and PANOVSKA, I. (2022). Impacts of COVID-19 local spread and Google search trend on the us stock market. *Phys. A, Stat. Mech. Appl.* **589** 126423.

DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. MR0520238

DOBRA, A., LENKOSKI, A. and RODRIGUEZ, A. (2011). Bayesian inference for general Gaussian graphical models with application to multivariate lattice data. *J. Amer. Statist. Assoc.* **106** 1418–1433. MR2896846 https://doi.org/10.1198/jasa.2011.tm10465

DOUCET, A., PITT, M. K., DELIGIANNIDIS, G. and KOHN, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102** 295–313. MR3371005 https://doi.org/10.1093/biomet/asu075

DUTTA, A., BOURI, E. and NOOR, M. H. (2021). Climate bond, stock, gold, and oil markets: Dynamic correlations and hedging analyses during the COVID-19 outbreak. *Resour. Policy* **74** 102265.

ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR0666121 https://doi.org/10.2307/1912773

ENGLE, R. F. and BOLLERSLEV, T. (1986). Modelling the persistence of conditional variances. *Econometric Rev.* **5** 1–87. With comments and a reply by the authors. MR0876792 https://doi.org/10.1080/07474938608800095

FRANZOLINI, B., BESKOS, A., DE IORIO, M., POKLEWSKI KOZIELL, W. and GRZESZKIEWICZ, K. (2024). Supplement to "Change point detection in dynamic Gaussian graphical models: The impact of COVID-19 pandemic on the U.S. stock market." https://doi.org/10.1214/23-AOAS1801SUPPA, https://doi.org/10.1214/23-AOAS1801SUPPB

FREEMAN, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry* 35–41.

GIBBERD, A. J. and NELSON, J. D. B. (2017). Regularized estimation of piecewise constant Gaussian graphical models: The group-fused graphical Lasso. *J. Comput. Graph. Statist.* **26** 623–634. MR3698672 https://doi.org/10.1080/10618600.2017.1302340

GIBBERD, A. J. and ROY, S. (2017). Multiple changepoint estimation in high-dimensional gaussian graphical models. arXiv preprint, arXiv:1712.05786.

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

GUIDOLIN, M., LA CARA, D. and MARCELLINO, M. G. (2021). Boosting the forecasting power of conditional heteroskedasticity models to account for COVID-19 outbreaks. BAFFI CAREFIN Centre Research Paper (2021-169).

HAAS, M., MITTNIK, S. and PAOLELLA, M. S. (2004). A new approach to Markov-switching GARCH models. *J. Financ. Econom.* **2** 493–530.

HALLAC, D., PARK, Y., BOYD, S. and LESKOVEC, J. (2017). Network inference via the time-varying graphical lasso. In *Proceedings of the* 23*rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 205–213.

HIGHAM, N. J. (2002). Computing the nearest correlation matrix—a problem from finance. *IMA J. Numer. Anal.* **22** 329–343. MR1918653 https://doi.org/10.1093/imanum/22.3.329

HLÁVKA, Z., HUŠKOVÁ, M. and MEINTANIS, S. G. (2020). Change-point methods for multivariate time-series: Paired vectorial observations. *Statist. Papers* **61** 1351–1383. MR4127478 https://doi.org/10.1007/s00362-020-01175-3

HULL, J. and WHITE, A. (1987). The pricing of options on assets with stochastic volatilities. *J. Finance* **42** 281–300.

JASRA, A., STEPHENS, D. A., DOUCET, A. and TSAGARIS, T. (2011). Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. *Scand. J. Stat.* **38** 1–22. MR2760137 https://doi.org/10.1111/j.1467-9469.2010.00723.x

JUST, M. and ECHAUST, K. (2020). Stock market returns, volatility, correlation and liquidity during the COVID-19 crisis: Evidence from the Markov switching approach. *Finance Res. Lett.* **37** 101775.

KARAGIANNIS, G. and ANDRIEU, C. (2013). Annealed importance sampling reversible jump MCMC algorithms. *J. Comput. Graph. Statist.* **22** 623–648. MR3173734 https://doi.org/10.1080/10618600.2013.805651

KESHAVARZ, H., MICHAILIDIS, G. and ATCHADÉ, Y. (2020). Sequential change-point detection in high-dimensional Gaussian graphical models. *J. Mach. Learn. Res.* **21** Paper No. 82, 57. MR4119150

KOLAR, M. and XING, E. P. (2012). Estimating networks with jumps. *Electron. J. Stat.* **6** 2069–2106. MR3020257 https://doi.org/10.1214/12-EJS739

KOTKATVUORI-ÖRNBERG, J., NIKKINEN, J. and ÄIJÖ, J. (2013). Stock market correlations during the financial crisis of 2008–2009: Evidence from 50 equity markets. *Int. Rev. Financ. Anal.* **28** 70–78.

LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. The Clarendon Press, Oxford Univ. Press, New York. Oxford Science Publications. MR1419991

LEDAY, G. G. R. and RICHARDSON, S. (2019). Fast Bayesian inference in large Gaussian graphical models. *Biometrics* **75** 1288–1298. MR4041830 https://doi.org/10.1111/biom.13064

LEE, H.-T. (2010). Regime switching correlation hedging. *J. Bank. Financ.* **34** 2728–2741.

LI, Y., CRAIG, B. A. and BHADRA, A. (2019). The graphical horseshoe estimator for inverse covariance matrices. *J. Comput. Graph. Statist.* **28** 747–757. MR4007755 https://doi.org/10.1080/10618600.2019.1575744

LIEN, D., TSE, Y. K. and TSUI, A. K. (2002). Evaluating the hedging performance of the constant-correlation GARCH model. *Appl. Financ. Econ.* **12** 791–798.

LINDSTEN, F., JORDAN, M. I. and SCHÖN, T. B. (2014). Particle Gibbs with ancestor sampling. *J. Mach. Learn. Res.* **15** 2145–2184. MR3231604

LIU, B., ZHANG, X. and LIU, Y. (2021). Simultaneous change point inference and structure recovery for high dimensional Gaussian graphical models. *J. Mach. Learn. Res.* **22** Paper No. [274], 62. MR4353053

LLOPIS, F. P., KANTAS, N., BESKOS, A. and JASRA, A. (2018). Particle filtering for stochastic Navier–Stokes signal observed with linear additive noise. *SIAM J. Sci. Comput.* **40** A1544–A1565. MR3807951 https://doi.org/10.1137/17M1151900

MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363 https://doi.org/10.1214/009053606000000281

MICHIS, A. A. (2022). Multiscale partial correlation clustering of stock market returns. *J. Risk Financ. Manag.* **15** 24.

MOHAMMADI, R., WIT, E. and DOBRA, A. (2022). BDgraph: Bayesian structure learning in graphical models using birth-death MCMC. R package version 2.72.

MOLINARI, M., CREMASCHI, A., DE IORIO, M., CHATURVEDI, N., HUGHES, A. and TILLIN, T. (2022a). Bayesian dynamic network modelling: An application to metabolic associations in cardiovascular disease. arXiv preprint, arXiv:2207.13418.

MOLINARI, M., CREMASCHI, A., DE IORIO, M., CHATURVEDI, N., HUGHES, A. D. and TILLIN, T. (2022b). Bayesian nonparametric modelling of multiple graphs with an application to ethnic metabolic differences. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **71** 1181–1204. MR4511106 https://doi.org/10.1111/rssc.12570

MURRAY, I., GHAHRAMANI, Z. and MACKAY, D. J. C. (2006). MCMC for doubly-intractable distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, *UAI'*06.

NEWTON, M. A., NOUEIRY, A., SARKAR, D. and AHLQUIST, P. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* **5** 155–176.

NILSSON, O. (2016). On stochastic volatility models as an alternative to GARCH type models. Dept. Statistics, Uppsala Univ., Sweden.

PERSING, A., JASRA, A., BESKOS, A., BALDING, D. and DE IORIO, M. (2015). A simulation approach for change-points on phylogenetic trees. *J. Comput. Biol.* **22** 10–24.

PETERSON, C., STINGO, F. C. and VANNUCCI, M. (2015). Bayesian inference of multiple Gaussian graphical models. *J. Amer. Statist. Assoc.* **110** 159–174. MR3338494 https://doi.org/10.1080/01621459.2014.896806

PITT, M. K., SILVA, R. D. S., GIORDANI, P. and KOHN, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *J. Econometrics* **171** 134–151. MR2991856 https://doi.org/10.1016/j.jeconom.2012.06.004

PRADO, R. and WEST, M. (2010). *Time Series*: *Modeling*, *Computation*, *and Inference*. CRC Press/CRC.

ROVERATO, A. (2002). Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scand. J. Stat.* **29** 391–411. MR1925566 https://doi.org/10.1111/1467-9469.00297

ROY, S., ATCHADÉ, Y. and MICHAILIDIS, G. (2017). Change point estimation in high dimensional Markov random-field models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1187–1206. MR3689314 https://doi.org/10.1111/rssb.12205

RUZAYQAT, H., ER-RAIY, A., BESKOS, A., CRISAN, D., JASRA, A. and KANTAS, N. (2022). A lagged particle filter for stable filtering of certain high-dimensional state-space models. *SIAM/ASA J. Uncertain. Quantificat.* **10** 1130–1161. MR4490319 https://doi.org/10.1137/21M1450392

SAKURAI, Y. and KUROSAKI, T. (2020). How has the relationship between oil and the us stock market changed after the Covid-19 crisis? *Finance Res. Lett.* **37** 101773.

SCHWALLER, L. and ROBIN, S. (2017). Exact Bayesian inference for off-line change-point detection in tree-structured graphical models. *Stat. Comput.* **27** 1331–1345. MR3647100 https://doi.org/10.1007/s11222-016-9689-3

SHERLOCK, C., THIERY, A. H., ROBERTS, G. O. and ROSENTHAL, J. S. (2015). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.* **43** 238–275. MR3285606 https://doi.org/10.1214/14-AOS1278

SILVENNOINEN, A. and TERÄSVIRTA, T. (2009). Multivariate GARCH models. In *Handbook of Financial Time Series* 201–229. Springer.

SO, M. K. P., LAM, K. and LI, W. K. (1998). A stochastic volatility model with Markov switching. *J. Bus. Econom. Statist.* **16** 244–253. MR1615375 https://doi.org/10.2307/1392580

TAYLOR, S. J. (1982). Financial returns modelled by the product of two stochastic processes—a study of the daily sugar prices 1961–75. *Time Ser. Anal. Theory Pract.* **1** 203–226.

THAMPANYA, N., NASIR, M. A. and HUYNH, T. L. D. (2020). Asymmetric correlation and hedging effectiveness of gold & cryptocurrencies: From pre-industrial to the 4th industrial revolution. *Technol. Forecast. Soc. Change* **159** 120195.

UHLER, C., LENKOSKI, A. and RICHARDS, D. (2018). Exact formulas for the normalizing constants of Wishart distributions for graphical models. *Ann. Statist.* **46** 90–118. MR3766947 https://doi.org/10.1214/17-AOS1543

VAN LEEUWEN, P. J., CRISAN, D., LANG, O. and POTTHAST, R. (2021). Bayesian inference for fluid dynamics: A case study for the stochastic rotating shallow water model. arXiv preprint, arXiv:2112.15216.

VAN DEN BOOM, W., BESKOS, A. and DE IORIO, M. (2022). The G-Wishart weighted proposal algorithm: Efficient posterior computation for Gaussian graphical models. *J. Comput. Graph. Statist.* **31** 1215–1224. MR4513382 https://doi.org/10.1080/10618600.2022.2050250

VAN DEN BOOM, W., DE IORIO, M. and BESKOS, A. (2022). Bayesian learning of graph substructures. arXiv preprint, arXiv:2203.11664.

VAN DEN BOOM, W., JASRA, A., DE IORIO, M., BESKOS, A. and ERIKSSON, J. G. (2022). Unbiased approximation of posteriors via coupled particle Markov chain Monte Carlo. *Stat. Comput.* **32** Paper No. 36, 19. MR4414445 https://doi.org/10.1007/s11222-022-10093-3

VON FURSTENBERG, G. M., JEON, B. N., MANKIW, N. G. and SHILLER, R. J. (1989). International stock price movements: Links and messages. *Brookings Pap. Econ. Act.* **1989** 125–179.

WANG, H. (2010). Sparse seemingly unrelated regression modelling: Applications in finance and econometrics. *Comput. Statist. Data Anal.* **54** 2866–2877. MR2720481 https://doi.org/10.1016/j.csda.2010.03.028

WANG, H. and WEST, M. (2009). Bayesian analysis of matrix normal graphical models. *Biometrika* **96** 821–834. MR2564493 https://doi.org/10.1093/biomet/asp049

WARNICK, R., GUINDANI, M., ERHARDT, E., ALLEN, E., CALHOUN, V. and VANNUCCI, M. (2018). A Bayesian approach for estimating dynamic functional network connectivity in fMRI data. *J. Amer. Statist. Assoc.* **113** 134–151. MR3803445 https://doi.org/10.1080/01621459.2017.1379404

WATTS, D. J. and STROGATZ, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* **393** 440–442.

WEST, M. and HARRISON, J. (2006). *Bayesian Forecasting and Dynamic Models*. Springer Science & Business Media. MR1020301 https://doi.org/10.1007/978-1-4757-9365-9

WIGGINS, J. B. (1987). Option values under stochastic volatility: Theory and empirical estimates. *J. Financ. Econ.* **19** 351–372.

WILLIAMS, D. R. (2021). Bayesian estimation for Gaussian graphical models: Structure learning, predictability, and network comparisons. *Multivar. Behav. Res.* **56** 336–352. https://doi.org/10.1080/00273171.2021.1894412

YANG, J. and PENG, J. (2020). Estimating time-varying graphical models. *J. Comput. Graph. Statist.* **29** 191–202. MR4085874 https://doi.org/10.1080/10618600.2019.1647848

YANG, S., LU, Z., SHEN, X., WONKA, P. and YE, J. (2015). Fused multiple graphical lasso. *SIAM J. Optim.* **25** 916–943. MR3343365 https://doi.org/10.1137/130936397

YILDIRIM, S., SINGH, S. S. and DOUCET, A. (2013). An online expectation-maximization algorithm for changepoint models. *J. Comput. Graph. Statist.* **22** 906–926. MR3173749 https://doi.org/10.1080/10618600.2012.674653

YIN, J. and LI, H. (2011). A sparse conditional Gaussian graphical model for analysis of genetical genomics data. *Ann. Appl. Stat.* **5** 2630–2650. MR2907129 https://doi.org/10.1214/11-AOAS494

YOUSFI, M., ZAIED, Y. B., CHEIKH, N. B., LAHOUEL, B. B. and BOUZGARROU, H. (2021). Effects of the COVID-19 pandemic on the us stock market and uncertainty: A comparative assessment between the first and second waves. *Technol. Forecast. Soc. Change* **167** 120710.

ZHOU, S., LAFFERTY, J. and WASSERMAN, L. (2010). Time varying undirected graphs. *Mach. Learn.* **80** 295–319. MR3108169 https://doi.org/10.1007/s10994-010-5180-0

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

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 585–599
https://doi.org/10.1214/23-AOAS1802

A NOVEL ESTIMATOR OF EARTH'S CURVATURE (ALLOWING FOR INFERENCE AS WELL)

BY DAVID R. BELL[1,a], OLIVIER LEDOIT[2,b] AND MICHAEL WOLF[2,c]

[1]*Idea Farm Ventures,* [a]*david@ideafarmventures.com*
[2]*Department of Economics, University of Zurich,* [b]*olivier.ledoit@econ.uzh.ch,* [c]*michael.wolf@econ.uzh.ch*

This paper estimates the curvature of the Earth, defined as one over its radius, without relying on physical measurements. The orthodox model states that the Earth is (nearly) spherical with a curvature of $\pi/20{,}000$ km. By contrast, the heterodox flat-Earth model stipulates a curvature of zero. Abstracting from the well-worn arguments for and against both models, rebuttals and counter-rebuttals ad infinitum, we propose a novel statistical methodology based on verifiable flight times along regularly scheduled commercial airline routes; this methodology allows for both estimating and making inference for Earth's curvature. In particular, a formal hypothesis test resolutely rejects the flat-Earth model, whereas it does not reject the orthodox spherical-Earth model.

REFERENCES

BELL, D. R, LEDOIT, O. and WOLF, M. (2024). Supplement to "A novel estimator of Earth's curvature (Allowing for inference as well)." https://doi.org/10.1214/23-AOAS1802SUPPA, https://doi.org/10.1214/23-AOAS1802SUPPB

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

GREENE, W. H. (2008). *Econometric Analysis*, sixth ed. Pearson, Upper Saddle River, NJ.

KUHN, T. S. (1962). *The Structure of Scientific Revolutions*. Univ. Chicago Press, Chicago, IL.

KUZII, O. and ROVENCHAK, A. (2019). What the gravitation of a flat Earth would look like and why thus the Earth is not actually flat. *Eur. J. Phys*. **40** 035008.

LEDSOM, A. (2022). Why Is Your Flight Slower Today Than Ever Before? Forbes, (published on 08-December-2022). https://www.forbes.com/sites/alexledsom/.

POPPER, K. R. (1959). *The Logic of Scientific Discovery*. Hutchinson, London. MR0107593

ROWBOTHAM, S. B. (1881). *Zetetic Astronomy*: *Earth Not a Globe*, third ed.

SNYDER, P. J. (1987). Map Projections – A Working Manual. U.S. Geological Survey #1395.

Key words and phrases. Flat Earth, flight times, nonlinear least squares, trigonometry.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 600–625
https://doi.org/10.1214/23-AOAS1803

ONLINE MONITORING OF AIR QUALITY USING PCA-BASED SEQUENTIAL LEARNING

BY XIULIN XIE[1,a], NICOLE QIAN[2,b] AND PEIHUA QIU[3,c]

[1]*Department of Statistics, Florida State University,* [a]*xx23d@fsu.edu*
[2]*Department of Engineering Sciences & Applied Mathematics, Northwestern University,* [b]*nicoleqian2027@u.northwestern.edu*
[3]*Department of Biostatistics, University of Florida,* [c]*pqiu@ufl.edu*

Air pollution surveillance is critically important for public health. One air pollutant, ozone, is extremely challenging to analyze properly, as it is a secondary pollutant caused by complex chemical reactions in the air and does not emit directly into the atmosphere. Numerous environmental studies confirm that ozone concentration levels are associated with meteorological conditions, and long-term exposure to high ozone concentration levels is associated with the incidence of many diseases, including asthma, respiratory, and cardiovascular diseases. Thus, it is important to develop an air pollution surveillance system to collect both air pollution and meteorological data and monitor the data continuously over time. To this end, statistical process control (SPC) charts provide a major statistical tool. But most existing SPC charts are designed for cases when the in-control (IC) process observations at different times are assumed to be independent and identically distributed. The air pollution and meteorological data would not satisfy these conditions due to serial data correlation, high dimensionality, seasonality, and other complex data structure. Motivated by an application to monitor the ground ozone concentration levels in the Houston–Galveston–Brazoria (HGB) area, we developed a new process monitoring method using principal component analysis and sequential learning. The new method can accommodate high dimensionality, time-varying IC process distribution, serial data correlation, and nonparametric data distribution. It is shown to be a reliable analytic tool for online monitoring of air quality.

REFERENCES

ABDUL-WAHAB, S. A., BAKHEIT, C. S. and AL-ALAWI, S. M. (2005). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environ. Model. Softw.* **20** 1263–1271.

ALAVA, J. J. and SINGH, G. G. (2022). Changing air pollution and CO2 emissions during the COVID-19 pandemic: Lesson learned and future equity concerns of post-COVID recovery. *Environ. Sci. Policy* **130** 1–8.

ALTMAN, N. S. (1990). Kernel smoothing of data with correlated errors. *J. Amer. Statist. Assoc.* **85** 749–759. MR1138355

APLEY, D. W. and TSUNG, F. (2002). The autoregressive t2 chart for monitoring univariate autocorrelated processes. *J. Qual. Technol.* **34** 80–96.

BARUA, S. and NATH, S. D. (2021). The impact of COVID-19 on air pollution: Evidence from global data. *J. Clean. Prod.* **298** 126755. https://doi.org/10.1016/j.jclepro.2021.126755

BERAN, J. (1992). Statistical methods for data with long-range dependence. *Statist. Sci.* **7** 404–416.

CAPIZZI, G. and MASAROTTO, G. (2008). Practical design of generalized likelihood ratio control charts for autocorrelated data. *Technometrics* **50** 357–370. MR2528658 https://doi.org/10.1198/004017008000000280

CAPIZZI, G. and MASAROTTO, G. (2011). A least angle regression control chart for multidimensional data. *Technometrics* **53** 285–296. MR2867502 https://doi.org/10.1198/TECH.2011.10027

CAREY, I. M., ATKINSON, R. W., KENT, A. J., VAN STAA, T., COOK, D. G. and ANDERSON, H. R. (2013). Mortality associations with long-term exposure to outdoor air pollution in a national English cohort. *Am. J. Respir. Crit. Care Med.* **187** 1226–1233.

Key words and phrases. Air pollution surveillance, dynamic processes, high-dimensional data, principal component analysis, seasonality, self-starting charts.

CHATTERJEE, S. and QIU, P. (2009). Distribution-free cumulative sum control charts using bootstrap-based control limits. *Ann. Appl. Stat.* **3** 349–369. MR2668711 https://doi.org/10.1214/08-AOAS197

CHICKEN, E., PIGNATIELLO, J. J. and SIMPSON, J. R. (2009). Statistical process monitoring of nonlinear profiles using wavelets. *J. Qual. Technol.* **41** 198–212.

CROSIER, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics* **30** 291–303. MR0959530 https://doi.org/10.2307/1270083

DE BRABANTER, K., DE BRABANTER, J., SUYKENS, J. A. K. and DE MOOR, B. (2011). Kernel regression in the presence of correlated errors. *J. Mach. Learn. Res.* **12** 1955–1976. MR2819023

DE KETELAERE, B., HUBERT, M. and SCHMITT, E. (2015). Overview of PCA-based statistical process-monitoring methods for time-dependent, high-dimensional data. *J. Qual. Technol.* **47** 318–335.

DONG, Y. and QIN, S. J. (2018). A novel dynamic PCA algorithm for dynamic data modeling and process monitoring. *J. Process Control* **67** 1–11.

DRAXLER, R. R. (2000). Meteorological factors of ozone predictability at Houston, Texas. *J. Air Waste Manage. Assoc.* **50** 259–271.

ENVIRONMENTAL PROTECTION AGENCY (1999). *Guideline for Developing an Ozone Forecasting Program*. Environmental Protection Agency, Washington.

EPANECHNIKOV, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory Probab. Appl.* **14** 153–158. MR0250422

FERRER, A. (2007). Multivariate statistical process control based on principal component analysis (MSPC-PCA): Some reflections and a case study in an autobody assembly process. *Qual. Eng.* **19** 311–325.

GORAI, A. K., TULURI, F., TCHOUNWOU, P. B. and AMBINAKUDIGE, S. (2015). Influence of local meteorology and NO2 conditions on ground-level ozone concentrations in the eastern part of Texas, USA. *Air Qual. Atmos. Health* **8** 81–96.

HAWKINS, D. M. (1987). Self-starting cusum charts for location and scale. *J. R. Stat. Soc., Ser. D, Stat.* **36** 299–316.

HAWKINS, D. M., QIU, P. and KANG, C. W. (2003). The changepoint model for statistical process control. *J. Qual. Technol.* **35** 355–366.

HEALTH EFFECTS INSTITUTE (2019). *State of Global Air* 2019. Health Effects Institute, Boston, MA.

HOTELLING, H. (1947). Multivariate quality control. In *Techniques of Statistical Analysis* (C. Eisenhart, M. Hastay and W. A. Wallis, eds.) 111–184. McGraw Hill.

JACKSON, J. E. (1991). *A User's Guide to Principal Components*. Wiley, New York.

JACOB, D. J. and WINNER, D. A. (2009). Effect of climate change on air quality. *Atmos. Environ.* **43** 51–63.

JENKIN, M. E. and CLEMITSHAW, K. C. (2000). Ozone and other secondary photochemical pollutants: Chemical processes governing their formation in the planetary boundary layer. *Atmos. Environ.* **34** 2499–2527.

JOHNSON, R. A. and WICHERN, D. W. (2007). *Applied Multivariate Statistical Analysis*, 6th ed. Pearson Prentice Hall, Upper Saddle River, NJ. MR2372475

KNOTH, S., SALEH, N. A., MAHMOUD, M. A., WOODALL, W. H. and TERCERO-GÓMEZ, V. G. (2023). A critique of a variety of "memory-based" process monitoring methods. *J. Qual. Technol.* **55** 18–42.

KNOTH, S. and SCHMID, W. (2004). Control charts for time series: A review. In *Frontiers in Statistical Quality Control* 7. *Front. Stat. Qual. Control* 210–236. Physica, Heidelberg. MR2071564

KNOTH, S., TERCERO-GÓMEZ, V. G., KHAKIFIROOZ, M. and WOODALL, W. H. (2021). The impracticality of homogeneously weighted moving average and progressive mean control chart approaches. *Qual. Reliab. Eng. Int.* **37** 3779–3794.

KOURTI, T. and MACGREGOR, J. F. (1996). Multivariate SPC methods for process and product monitoring. *J. Qual. Technol.* **28** 409–428.

KU, W., STORER, R. H. and GEORGAKIS, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemom. Intell. Lab. Syst.* **30** 179–196.

LENNOX, B., MONTAGUE, G. A., HIDEN, H. G., KORNFELD, G. and GOULDING, P. R. (2001). Process monitoring of an industrial fed-batch fermentation. *Biotechnol. Bioeng.* **74** 125–135. https://doi.org/10.1002/bit.1102

LI, G., QIN, S. J. and ZHOU, D. (2014). A new method of dynamic latent-variable modeling for process monitoring. *IEEE Trans. Ind. Electron.* **61** 6438–6445.

LI, W., XIANG, D., TSUNG, F. and PU, X. (2020). A diagnostic procedure for high-dimensional data streams via missed discovery rate control. *Technometrics* **62** 84–100. MR4058601 https://doi.org/10.1080/00401706.2019.1575284

LIU, Y., ZHOU, Y. and LU, J. (2020). Exploring the relationship between air pollution and meteorological conditions in China under environmental governance. *Sci. Rep.* **10** 14518.

LOWRY, C. A., WOODALL, W. H., CHAMP, C. W. and RIGDON, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics* **34** 46–53.

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

MONTGOMERY, D. C. (2012). *Introduction to Statistical Quality Control*. Wiley, New York.

NOOROSSANA, R., SAGHAEI, A. and AMIRI, A. (2011). *Statistical Analysis of Profile Monitoring*. Chapman Hall/CRC, Boca Raton, FL.

OPSOMER, J., WANG, Y. and YANG, Y. (2001). Nonparametric regression with correlated errors. *Statist. Sci.* **16** 134–153. MR1861070 https://doi.org/10.1214/ss/1009213287

ORDÓÑEZ, C., MATHIS, H., FURGER, M., HENNE, S., HÜGLIN, C., STAEHELIN, J. and PRÉVÔT, A. S. H. (2005). Changes of daily surface ozone maxima in Switzerland in all seasons from 1992 to 2002 and discussion of summer 2003. *Atmos. Chem. Phys.* **5** 1187–1203.

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

PSARAKIS, S. and PAPALEONIDA, G. E. A. (2007). SPC procedures for monitoring autocorrelated processes. *Qual. Technol. Quant. Manag.* **4** 501–540. MR2422803 https://doi.org/10.1080/16843703.2007.11673168

QIU, P. (2014). *Introduction to Statistical Process Control*. Chapman Hall/CRC, Boca Raton, FL.

QIU, P. (2018). Some perspectives on nonparametric statistical process control. *J. Qual. Technol.* **50** 49–65.

QIU, P., LI, W. and LI, J. (2020). A new process control chart for monitoring short-range serially correlated data. *Technometrics* **62** 71–83. MR4058600 https://doi.org/10.1080/00401706.2018.1562988

QIU, P. and XIANG, D. (2014). Univariate dynamic screening system: An approach for identifying individuals with irregular longitudinal behavior. *Technometrics* **56** 248–260. MR3207851 https://doi.org/10.1080/00401706.2013.822423

QIU, P. and XIE, X. (2022). Transparent sequential learning for statistical process control of serially correlated data. *Technometrics* **64** 487–501. MR4506615 https://doi.org/10.1080/00401706.2021.1929493

QIU, P., ZOU, C. and WANG, Z. (2010). Nonparametric profile monitoring by mixed effects modeling. *Technometrics* **52** 265–277. MR2723706 https://doi.org/10.1198/TECH.2010.08188

ROBERTS, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics* **1** 239–250.

SEXTON, K. and LINDER, S. H. (2015). Houston's novel strategy to control hazardous air pollutants: A case study in policy innovation and political stalemate. *Environ. Health Insights* **9** 1–12. https://doi.org/10.4137/EHI.S15670

SHEWHART, W. A. (1931). *Economic Control of Quality of Manufactured Product*. D. Van Nostrand Company, New York.

STATHEROPOULOS, M., VASSILIADIS, N. and PAPPA, A. (1998). Principal component and canonical correlation analysis for examining air pollution and meteorological data. *Atmos. Environ.* **32** 1087–1095.

SUN, W., PALAZOGLU, A., SINGH, A., ZHANG, H., WANG, Q., ZHAO, Z. and CAO, D. (2015). Prediction of surface ozone episodes using clusters based generalized linear mixed effects models in Houston–Galveston–Brazoria area, Texas. *Atmos. Pollut. Res.* **6** 245–253.

TARTAKOVSKY, A. G., ROZOVSKII, B. L., BLAŽEK, R. B. and KIM, H. (2006). Detection of intrusions in information systems by sequential change-point methods. *Stat. Methodol.* **3** 252–293. MR2240956 https://doi.org/10.1016/j.stamet.2005.05.003

TSUNG, F. (2000). Statistical monitoring and diagnosis of automatic controlled processes using dynamic PCA. *Int. J. Prod. Res.* **38** 625–637.

VANHATALO, E. and KULAHCI, M. (2016). Impact of autocorrelation on principal components and their use in statistical process control. *Qual. Reliab. Eng. Int.* **32** 1483–1500.

VENTER, Z., AUNAN, K., CHOWDHURY, S. and LELIEVELD, J. (2020). COVID-19 lockdowns cause global air pollution declines. *Proc. Natl. Acad. Sci. USA* **117** 18984–18990.

WANG, K. and JIANG, W. (2009). High-dimensional process monitoring and fault isolation via variable selection. *J. Qual. Technol.* **41** 247–258.

WANG, X., KRUGER, U. and IRWIN, G. W. (2005). Process monitoring approach using fast moving window PCA. *Ind. Eng. Chem. Res.* **44** 5691–5702.

WENG, J., ZHANG, Y. and HWANG, W.-S. (2003). Candid covariance-free incremental principal component analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **25** 1034–1040.

WORLD HEALTH ORGANIZATION (1976). *Photochemical Oxidants*: *Environmental Health Criteria* 7. World Health Organization, Geneva.

XIANG, D., QIU, P. and PU, X. (2013). Nonparametric regression analysis of multivariate longitudinal data. *Statist. Sinica* **23** 769–789. MR3086655

XIANG, D., QIU, P., WANG, D. and LI, W. (2022). Reliable post-signal fault diagnosis for correlated high-dimensional data streams. *Technometrics* **64** 323–334. MR4457326 https://doi.org/10.1080/00401706.2021.1979100

XIE, X., QIAN, N. and QIU, P. (2024). Supplement to "Online monitoring of air quality using PCA-based sequential learning." https://doi.org/10.1214/23-AOAS1803SUPP

XIE, X. and QIU, P. (2023). Control charts for dynamic process monitoring with an application to air pollution surveillance. *Ann. Appl. Stat.* **17** 47–66. MR4539021 https://doi.org/10.1214/22-aoas1615

YANG, K. and QIU, P. (2018). Spatiotemporal incidence rate data analysis by nonparametric regression. *Stat. Med.* **37** 2094–2107. MR3802936 https://doi.org/10.1002/sim.7622

ZHANG, K. and FAN, W. (2008). Forecasting skewed biased stochastic ozone days: Analyses, solutions and beyond. *Knowl. Inf. Syst.* **14** 299–326.

ZHAO, X., ZHANG, X., XU, X., XU, J., MENG, W. and PU, W. (2009). Seasonal and diurnal variations of ambient PM2.5 concentration in urban and rural environments in Beijing. *Atmos. Environ.* **43** 2893–2900.

ZOU, C. and QIU, P. (2009). Multivariate statistical process control using LASSO. *J. Amer. Statist. Assoc.* **104** 1586–1596. MR2750580 https://doi.org/10.1198/jasa.2009.tm08128

ZOU, C., WANG, Z., JIANG, W. and ZI, X. (2015). An efficient online monitoring method for high-dimensional data streams. *Technometrics* **57** 374–387. MR3384952 https://doi.org/10.1080/00401706.2014.940089

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 626–641
https://doi.org/10.1214/23-AOAS1804

TENSOR MIXTURE DISCRIMINANT ANALYSIS WITH APPLICATIONS TO SENSOR ARRAY DATA ANALYSIS

BY XUESONG HOU[1,a], QING MAI[2,c] AND HUI ZOU[1,b]

[1]*School of Statistics, University of Minnesota,*[a]*hou00034@umn.edu,* [b]*zouxx019@umn.edu*
[2]*Department of Statistics, Florida State University,*[c]*qmai@fsu.edu*

Sensor arrays are often used to identify chemicals by measuring properly chosen chemical interactions. Machine learning techniques are of vital importance to accurately recognize a chemical based on the sensor array measurements. However, sensor array data often take the form of matrices (i.e, two-way tensors), and the concentration levels may have a complex impact on the measurements. Hence, existing linear and/or vector classification methods may be inadequate for sensor array data. In this article we propose a novel tensor mixture discriminant analysis (TMDA) model carefully tailored for the classification of sensor array data. We model the distribution of each chemical by a mixture of tensor normal distributions. TMDA leverages the tensor structure for better estimation and prediction, while the mixed tensor normal component accounts for the possibly varying concentration levels. The TMDA model can also be viewed as an approximation of the potentially nonnormal measurements. An efficient expectation-maximization algorithm is developed to fit the TMDA model. The application of TMDA on two sensor array datasets demonstrates its superior performance to many popular competitors.

REFERENCES

BOSER, B. E., GUYON, I. M. and VAPNIK, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory. COLT* '92 144–152. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/130385.130401

BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32. https://doi.org/10.1023/A:1010933404324

CAI, B., ZHANG, J. and SUN, W. W. (2021). Heterogeneous tensor mixture models in high dimensions.

CLEMMENSEN, L., HASTIE, T., WITTEN, D. and ERSBØLL, B. (2011). Sparse discriminant analysis. *Technometrics* **53** 406–413. MR2850472 https://doi.org/10.1198/TECH.2011.08118

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

FRIEDMAN, J. H., HASTIE, T. and TIBSHIRANI, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22. https://doi.org/10.18637/jss.v033.i01

GUPTA, A. K. and NAGAR, D. K. (2000). *Matrix Variate Distributions. Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics* **104**. CRC Press/CRC, Boca Raton, FL. MR1738933

HASTIE, T. and TIBSHIRANI, R. (1996). Discriminant analysis by Gaussian mixtures. *J. Roy. Statist. Soc. Ser. B* **58** 155–176. MR1379236

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 https://doi.org/10.1007/978-0-387-84858-7

HOFF, P. D. (2011). Separable covariance arrays via the Tucker product, with applications to multivariate relational data. *Bayesian Anal.* **6** 179–196. MR2806238 https://doi.org/10.1214/11-BA606

HOU, X., MAI, Q. and ZOU, H. (2024). Supplement to "Tensor mixture discriminant analysis with applications to sensor array data analysis." https://doi.org/10.1214/23-AOAS1804SUPP

KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056 https://doi.org/10.1137/07070111X

KONG, H., TEOH, E. K., WANG, J. G. and VENKATESWARLU, R. (2005). Two-dimensional Fisher discriminant analysis: Forget about small sample size problem [face recognition applications]. In *Proceedings.*

Key words and phrases. Sensor arrays, tensor data classification, tensor normal mixtures.

(*ICASSP* '05). *IEEE International Conference on Acoustics*, *Speech*, *and Signal Processing* **2** ii/761–ii/764. https://doi.org/10.1109/ICASSP.2005.1415516

LAI, Z., XU, Y., YANG, J., TANG, J. and ZHANG, D. (2013). Sparse tensor discriminant analysis. *IEEE Trans*. *Image Process*. **22** 3904–3915. MR3105150 https://doi.org/10.1109/TIP.2013.2264678

LI, Q. and SCHONFELD, D. (2014). Multilinear discriminant analysis for higher-order tensor data classification. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **36** 2524–2537. https://doi.org/10.1109/TPAMI.2014.2342214

LI, X., XU, D. and ZHOU, H. (2018). Tucker tensor regression and neuroimaging analysis. *Stat*. *Biosci*. **10** 520–545.

LYU, X., SUN, W. W., WANG, Z., LIU, H., YANG, J. and CHENG, G. (2019). Tensor graphical model: Nonconvex optimization and statistical inference. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **42** 2024–2037.

MAI, Q., ZHANG, X., PAN, Y. and DENG, K. (2022). A doubly enhanced EM algorithm for model-based tensor clustering. *J*. *Amer*. *Statist*. *Assoc*. **117** 2120–2134. MR4528493 https://doi.org/10.1080/01621459.2021.1904959

MIN, K., MAI, Q. and ZHANG, X. (2022). Fast and separable estimation in high-dimensional tensor Gaussian graphical models. *J*. *Comput*. *Graph*. *Statist*. **31** 294–300. MR4387228 https://doi.org/10.1080/10618600.2021.1938086

MOLSTAD, A. J. and ROTHMAN, A. J. (2019). A penalized likelihood method for classification with matrix-valued predictors. *J*. *Comput*. *Graph*. *Statist*. **28** 11–22. MR3939368 https://doi.org/10.1080/10618600.2018.1476249

PAN, Y., MAI, Q. and ZHANG, X. (2019). Covariate-adjusted tensor classification in high dimensions. *J*. *Amer*. *Statist*. *Assoc*. **114** 1305–1319. MR4011781 https://doi.org/10.1080/01621459.2018.1497500

PERSAUD, K. and DODD, G. (1982). Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. *Nature* **299** 352–355. https://doi.org/10.1038/299352a0

SUSLICK, K., BAILEY, D., INGISON, C., JANZEN, M., KOSAL, M., MCNAMARA, W., RAKOW, N., SEN, A. and WEAVER, J. (2007). Seeing smells: Development of an optoelectronic nose. *Quím*. *Nova* **30** 677–681. https://doi.org/10.1590/S0100-40422007000300029

TAO, D., LI, X. and WU, X. (2007). General tensor discriminant analysis and Gabor features for gait recognition. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **29** 1700–1715. https://doi.org/10.1109/TPAMI.2007.1096

VERGARA, A., VEMBU, S., AYHAN, T., RYAN, M. A., HOMER, M. L. and HUERTA, R. (2012). Chemical gas sensor drift compensation using classifier ensembles. *Sens*. *Actuators B*, *Chem*. **166–167** 320–329. https://doi.org/10.1016/j.snb.2012.01.074

WITTEN, D. M. and TIBSHIRANI, R. (2011). Penalized classification using Fisher's linear discriminant. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **73** 753–772. MR2867457 https://doi.org/10.1111/j.1467-9868.2011.00783.x

YAN, S., XU, D., YANG, Q., ZHANG, L. and TANG, X. (2005). Discriminant analysis with tensor representation. In 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (*CVPR*'05) **1** 526–532. https://doi.org/10.1109/CVPR.2005.131

ZHANG, A. R., LUO, Y., RASKUTTI, G. and YUAN, M. (2020). ISLET: Fast and optimal low-rank tensor regression via importance sketching. *SIAM J*. *Math*. *Data Sci*. **2** 444–479. MR4106613 https://doi.org/10.1137/19M126476X

ZHONG, W. and SUSLICK, K. S. (2015). Matrix discriminant analysis with application to colorimetric sensor array data. *Technometrics* **57** 524–534. MR3425489 https://doi.org/10.1080/00401706.2014.965347

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 642–663
https://doi.org/10.1214/23-AOAS1805

BAYESIAN ADAPTIVE AND INTERPRETABLE FUNCTIONAL REGRESSION FOR EXPOSURE PROFILES

BY YUNAN GAO[a] AND DANIEL R. KOWAL[b]

Department of Statistics, Rice University, [a]*yunan.gao@rice.edu,* [b]*daniel.kowal@rice.edu*

Pollutant exposure during gestation is a known and adverse factor for birth and health outcomes. However, the links between prenatal air pollution exposures and educational outcomes are less clear, in particular, the critical windows of susceptibility during pregnancy. Using a large cohort of students in North Carolina, we study the link between prenatal daily $\mathrm{PM}_{2.5}$ exposure and fourth end-of-grade reading scores. We develop and apply a locally adaptive and highly scalable Bayesian regression model for scalar responses with functional and scalar predictors. The proposed model pairs a B-spline basis expansion with dynamic shrinkage priors to capture both smooth and rapidly-changing features in the regression surface. The model is accompanied by a new decision analysis approach for functional regression that extracts the critical windows of susceptibility and guides the model interpretations. These tools help to identify and address broad limitations with the interpretability of functional regression models. Simulation studies demonstrate more accurate point estimation, more precise uncertainty quantification, and far superior window selection than existing approaches. Leveraging the proposed modeling, computational, and decision analysis framework, we conclude that prenatal $\mathrm{PM}_{2.5}$ exposure during early and late pregnancy is most adverse for fourth end-of-grade reading scores.

REFERENCES

BASHIR, A., CARVALHO, C. M., HAHN, P. R. and JONES, M. B. (2019). Post-processing posteriors over precision matrices to produce sparse graph estimates. *Bayesian Anal.* **14** 1075–1090.

BERGER, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* 716–761.

BOSE, S., CHIU, Y.-H. M., HSU, H.-H. L., DI, Q., ROSA, M. J., LEE, A., KLOOG, I., WILSON, A., SCHWARTZ, J. et al. (2017). Prenatal nitrate exposure and childhood asthma. Influence of maternal prenatal stress and fetal sex. *Am. J. Respir. Crit. Care Med.* **196** 1396–1403.

BOYLE, E. M., POULSEN, G., FIELD, D. J., KURINCZUK, J. J., WOLKE, D., ALFIREVIC, Z. and QUIGLEY, M. A. (2012). Effects of gestational age at birth on health outcomes at 3 and 5 years of age: Population based cohort study. *BMJ* **344** e896. https://doi.org/10.1136/bmj.e896

BRAVO, M. A. and MIRANDA, M. L. (2021). Effects of accumulated environmental, social and host exposures on early childhood educational outcomes. *Environ. Res.* **198** 111241. https://doi.org/10.1016/j.envres.2021.111241

BROWN, P. J., VANNUCCI, M. and FEARN, T. (1998). Multivariate Bayesian variable selection and prediction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 627–641.

CARDOT, H., FERRATY, F. and SARDA, P. (1999). Functional linear model. *Statist. Probab. Lett.* **45** 11–22.

CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480.

CHEN, X.-K., WEN, S. W., FLEMING, N., DEMISSIE, K., RHOADS, G. G. and WALKER, M. (2007). Teenage pregnancy and adverse birth outcomes: A large population based retrospective cohort study. *Int. J. Epidemiol.* **36** 368–373.

CHENG, Y., YIN, J., YANG, L., XU, M., LU, X., HUANG, W., DAI, G. and SUN, G. (2023). Ambient air pollutants in the first trimester of pregnancy and birth defects: An observational study. *BMJ Open* **13** e063712.

Key words and phrases. Decision analysis, functional data analysis, nonparametric regression, shrinkage, spline.

Chiu, Y.-H. M., Hsu, H.-H. L., Coull, B. A., Bellinger, D. C., Kloog, I., Schwartz, J., Wright, R. O. and Wright, R. J. (2016). Prenatal particulate air pollution and neurodevelopment in urban children: Examining sensitive windows and sex-specific associations. *Environ. Int.* **87** 56–65. https://doi.org/10.1016/j.envint.2015.11.010

Dziak, J. J., Coffman, D. L., Reimherr, M., Petrovich, J., Li, R., Shiffman, S. and Shiyko, M. P. (2019). Scalar-on-function regression for predicting distal outcomes from intensively gathered longitudinal data: Interpretability for applied scientists. *Stat. Surv.* **13** 150.

Feldman, J. and Kowal, D. R. (2022). Bayesian data synthesis and the utility-risk trade-off for mixed epidemiological data. *Ann. Appl. Stat.* **16** 2577–2602. https://doi.org/10.1214/22-AOAS1604

Figueiredo, M. A. (2003). Adaptive sparseness for supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **25** 1150–1159.

Gao, Y. and Kowal, D. R. (2024). Supplement to "Bayesian adaptive and interpretable functional regression for exposure profiles." https://doi.org/10.1214/23-AOAS1805SUPPA, https://doi.org/10.1214/23-AOAS1805SUPPB

Gogoi, M. (2014). Association of maternal age and low socio-economic status of women on birth outcome. *Int. Res. J. Soc. Sci.* **3** 21–27.

Goisis, A., Remes, H., Barclay, K., Martikainen, P. and Myrskylä, M. (2017). Advanced maternal age and the risk of low birth weight and preterm delivery: A within-family analysis using Finnish population registers. *Amer. J. Epidemiol.* **186** 1219–1226. https://doi.org/10.1093/aje/kwx177

Goldenberg, R. L., Culhane, J. F., Iams, J. D. and Romero, R. (2008). Epidemiology and causes of preterm birth. *Lancet* **371** 75–84.

Griffin, J. E. and Brown, P. J. (2005). Alternative prior distributions for variable selection with very many more variables than observations Technical Report Univ. Warwick, Centre for Research in Statistical Methodology.

Grollemund, P.-M., Abraham, C., Baragatti, M. and Pudlo, P. (2019). Bayesian functional linear regression with sparse step functions. *Bayesian Anal.* **14** 111–135.

Guxens, M., Lubczyńska, M. J., Muetzel, R. L., Dalmau-Bueno, A., Jaddoe, V. W., Hoek, G., van der Lugt, A., Verhulst, F. C., White, T. et al. (2018). Air pollution exposure during fetal life, brain morphology, and cognitive function in school-age children. *Biol. Psychiatry* **84** 295–303.

Hahn, P. R. and Carvalho, C. M. (2015). Decoupling shrinkage and selection in Bayesian linear models: A posterior summary perspective. *J. Amer. Statist. Assoc.* **110** 435–448.

Hazlehurst, M. F., Carroll, K. N., Loftus, C. T., Szpiro, A. A., Moore, P. E., Kaufman, J. D., Kirwa, K., LeWinn, K. Z., Bush, N. R. et al. (2021). Maternal exposure to PM2. 5 during pregnancy and asthma risk in early childhood: Consideration of phases of fetal lung development. *Environ. Epidemiol.* **5**.

Huang, C., Jiang, Q., Su, W., Lv, F., Zeng, J., Huang, P., Liu, W., Lin, M., Li, X. et al. (2023). Age-specific effects on adverse pregnancy outcomes vary by maternal characteristics: A population-based retrospective study in Xiamen, China. *BMC Public Health* **23** 326.

James, G. M. (2002). Generalized linear models with functional predictors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 411–432.

James, G. M., Wang, J. and Zhu, J. (2009). Functional linear regression that's interpretable. *Ann. Statist.* **37** 2083–2108.

Kalkbrenner, A. E., Windham, G. C., Serre, M. L., Akita, Y., Wang, X., Hoffman, K., Thayer, B. P. and Daniels, J. L. (2015). Particulate matter exposure, prenatal and postnatal windows of susceptibility, and autism spectrum disorders. *Epidemiology* **26** 30–42.

Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Rev. Econ. Stud.* **65** 361–393.

Kloog, I., Melly, S. J., Ridgway, W. L., Coull, B. A. and Schwartz, J. (2012). Using new satellite based exposure methods to study the association between pregnancy PM2.5 exposure, premature birth and birth weight in Massachusetts. *Environ. Health* **11** 1–8.

Kowal, D. R. (2021). Fast, optimal, and targeted predictions using parameterized decision analysis. *J. Amer. Statist. Assoc.* 1–12.

Kowal, D. R. (2022a). Bayesian subset selection and variable importance for interpretable prediction and classification. *J. Mach. Learn. Res.* **23** 1–38.

Kowal, D. R. (2022b). Subset selection for linear mixed models. *Biometrics* **00** 1–15.

Kowal, D. R. and Bourgeois, D. C. (2020). Bayesian function-on-scalars regression for high-dimensional data. *J. Comput. Graph. Statist.* **29** 629–638.

Kowal, D. R., Bravo, M., Leong, H., Bui, A., Griffin, R. J., Ensor, K. B. and Miranda, M. L. (2021). Bayesian variable selection for understanding mixtures in environmental exposures. *Stat. Med.* **40** 4850–4871. https://doi.org/10.1002/sim.9099

KOWAL, D. R., MATTESON, D. S. and RUPPERT, D. (2019). Dynamic shrinkage processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 781–804.

LEE, A., HSU, H.-H. L., CHIU, Y.-H. M., BOSE, S., ROSA, M. J., KLOOG, I., WILSON, A., SCHWARTZ, J., COHEN, S. et al. (2018). Prenatal fine particulate exposure and early childhood asthma: Effect of maternal stress and fetal sex. *J. Allergy Clin. Immunol.* **141** 1880–1886.

LEON HSU, H.-H., MATHILDA CHIU, Y.-H., COULL, B. A., KLOOG, I., SCHWARTZ, J., LEE, A., WRIGHT, R. O. and WRIGHT, R. J. (2015). Prenatal particulate air pollution and asthma onset in urban children. Identifying sensitive windows and sex differences. *Am. J. Respir. Crit. Care Med.* **192** 1052–1059.

LEUNG, M., WEISSKOPF, M. G., LADEN, F., COULL, B. A., MODEST, A. M., HACKER, M. R., WYLIE, B. J., WEI, Y., SCHWARTZ, J. et al. (2022). Exposure to PM 2.5 during pregnancy and fetal growth in eastern Massachusetts, USA. *Environ. Health Perspect.* **130** 017004.

MARX, B. D. and EILERS, P. H. C. (1999). Generalized linear regression on sampled signals and curves: A P-spline approach. *Technometrics* **41** 1–13.

MCGUINN, L. A., BELLINGER, D. C., COLICINO, E., COULL, B. A., JUST, A. C., KLOOG, I., OSORIO-VALENCIA, E., SCHNAAS, L., WRIGHT, R. J. et al. (2020). Prenatal PM2.5 exposure and behavioral development in children from Mexico City. *Neurotoxicology* **81** 109–115.

MIROWSKY, J. and ROSS, C. E. (2002). Depression, parenthood, and age at first birth. *Soc. Sci. Med.* **54** 1281–1298.

MORK, D. and WILSON, A. (2022). Treed distributed lag nonlinear models. *Biostatistics* **23** 754–771.

MORRIS, J. S. (2015). Functional regression. *Annu. Rev. Stat. Appl.* **2** 321–359.

MORRIS, J. S., BROWN, P. J., HERRICK, R. C., BAGGERLY, K. A. and COOMBES, K. R. (2008). Bayesian analysis of mass spectrometry proteomic data using wavelet-based functional mixed models. *Biometrics* **64** 479–489.

MORRIS, J. S. and CARROLL, R. J. (2006). Wavelet-based functional mixed models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 179–199. https://doi.org/10.1111/j.1467-9868.2006.00539.x

MÜLLER, H.-G. and STADTMÜLLER, U. (2005). Generalized functional linear models. *Ann. Statist.* **33** 774–805.

OLESEN, A. W., WESTERGAARD, J. G. and OLSEN, J. (2003). Perinatal and maternal complications related to postterm delivery: A national register-based study, 1978–1993. *Am. J. Obstet. Gynecol.* **189** 222–227. https://doi.org/10.1067/mob.2003.446

PICHENY, V., SERVIEN, R. and VILLA-VIALANEIX, N. (2019). Interpretable sparse SIR for functional data. *Stat. Comput.* **29** 255–267.

PUELZ, D., HAHN, P. R. and CARVALHO, C. M. (2017). Variable selection in seemingly unrelated regressions with random predictors. *Bayesian Anal.* **12** 969–989.

RAHMAN, F., COULL, B. A., CARROLL, K. N., WILSON, A., JUST, A. C., KLOOG, I., ZHANG, X., WRIGHT, R. J. and CHIU, Y.-H. M. (2021). Prenatal PM2. 5 exposure and infant temperament at age 6 months: Sensitive windows and sex-specific associations. *Environ. Res.* 112583.

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

SADLER, T. W. (2022). *Langman's Medical Embryology*. Williams & Wilkins, Baltimore.

SCHWARTZ, J. (2000). The distributed lag between air pollution and daily deaths. *Epidemiology* **11** 320–326. https://doi.org/10.1097/00001648-200005000-00016

SRÁM, R. J., BINKOVÁ, B., DEJMEK, J. and BOBAK, M. (2005). Ambient air pollution and pregnancy outcomes: A review of the literature. *Environ. Health Perspect.* **113** 375–382. https://doi.org/10.1289/ehp.6362

STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Stat.* **42** 385–388.

SUADES-GONZÁLEZ, E., GASCON, M., GUXENS, M. and SUNYER, J. (2015). Air pollution and neuropsychological development: A review of the latest evidence. *Endocrinology* **156** 3473–3482. https://doi.org/10.1210/en.2015-1403

SUN, X., LIU, C., JI, H., LI, W., MIAO, M., YUAN, W., YUAN, Z., LIANG, H. and KAN, H. (2023). Prenatal exposure to ambient PM2. 5 and its chemical constituents and child intelligence quotient at 6 years of age. *Ecotoxicol. Environ. Saf.* **255** 114813.

TIBSHIRANI, R. J. and TAYLOR, J. (2011). The solution path of the generalized lasso. *Ann. Statist.* **39** 1335–1371.

WARREN, J., FUENTES, M., HERRING, A. and LANGLOIS, P. (2012). Spatial-temporal modeling of the association between air pollution exposure and preterm birth: Identifying critical windows of exposure. *Biometrics* **68** 1157–1167. https://doi.org/10.1111/j.1541-0420.2012.01774.x

WARREN, J. L., KONG, W., LUBEN, T. J. and CHANG, H. H. (2020). Critical window variable selection: Estimating the impact of air pollution on very preterm birth. *Biostatistics* **21** 790–806.

WILSON, A., CHIU, Y.-H. M., HSU, H.-H. L., WRIGHT, R. O., WRIGHT, R. J. and COULL, B. A. (2017). Bayesian distributed lag interaction models to identify perinatal windows of vulnerability in children's health. *Biostatistics* **18** 537–552.

XIONG, L., XU, Z., WANG, H., LIU, Z., XIE, D., WANG, A. and KONG, F. (2019). The association between ambient air pollution and birth defects in four cities in Hunan province, China, from 2014 to 2016. *Medicine* **98**.

ZHANG, X. and KRAMER, M. S. (2009). Variations in mortality and morbidity by gestational age among infants born at term. *J. Pediatr.* **154** 358–362.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 664–682
https://doi.org/10.1214/23-AOAS1806

QUANTIFYING REPLICABILITY OF MULTIPLE STUDIES IN A META-ANALYSIS

BY MENGLI XIAO[1,a], HAITAO CHU[2,3,b], JAMES S. HODGES[3,c] AND LIFENG LIN[4,d]

[1]*Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus,* [a]*mengli.xiao@cuanschutz.edu*

[2]*Statistical Research and Data Science Center, Pfizer Inc.,* [b]*chux0051@umn.edu*

[3]*Division of Biostatistics, University of Minnesota,* [c]*hodge003@umn.edu*

[4]*Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona,* [d]*lifenglin@arizona.edu*

For valid scientific discoveries, it is fundamental to evaluate whether research findings are replicable across different settings. While large-scale replication projects across broad research topics are not feasible, systematic reviews and meta-analyses (SRMAs) offer viable alternatives to assess replicability. Due to subjective inclusion and exclusion of studies, SRMAs may contain nonreplicable study findings. However, there is no consensus on rigorous methods to assess the replicability of SRMAs or to explore sources of nonreplicability. Nonreplicability is often misconceived as high heterogeneity. This article introduces a new measure, the externally standardized residuals from a leave-m-studies-out procedure, to quantify replicability. It not only measures the impact of nonreplicability from unknown sources on the conclusion of an SRMA but also differentiates nonreplicability from heterogeneity. A new test statistic for replicability is derived. We explore its asymptotic properties and use extensive simulations and real data to illustrate this measure's performance. We conclude that replicability should be routinely assessed for all SRMAs and recommend sensitivity analyses, once nonreplicable study results are identified in an SRMA.

REFERENCES

ANG, A. H.-S. and TANG, W. H. (1984). *Probability Concepts in Engineering: Emphasis on Applications in Civil & Environmental Engineering*. Wiley, Hoboken, NJ.

ARELLANO-VALLE, R. B. and GENTON, M. G. (2008). On the exact distribution of the maximum of absolutely continuous dependent random variables. *Statist. Probab. Lett.* **78** 27–35. MR2381270 https://doi.org/10.1016/j.spl.2007.04.021

BEGLEY, C. G. and IOANNIDIS, J. P. A. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. *Circ. Res.* **116** 116–126. https://doi.org/10.1161/CIRCRESAHA.114.303819

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

BORENSTEIN, M., HEDGES, L. V., HIGGINS, J. P. T. and ROTHSTEIN, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res. Synth. Methods* **1** 97–111. https://doi.org/10.1002/jrsm.12

BRAVER, S. L., THOEMMES, F. J. and ROSENTHAL, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspect. Psychol. Sci.* **9** 333–342. https://doi.org/10.1177/1745691614529796

CAMERER, C. F., DREBER, A., FORSELL, E., HO, T. H., HUBER, J., JOHANNESSON, M., KIRCHLER, M., ALMENBERG, J., ALTMEJD, A. et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science* **351** 1433–1436. https://doi.org/10.1126/science.aaf0918

CIPRIANI, A., HIGGINS, J. P. T., GEDDES, J. R. and SALANTI, G. (2013). Conceptual and technical challenges in network meta-analysis. *Ann. Intern. Med.* **159** 130–137. https://doi.org/10.7326/0003-4819-159-2-201307160-00008

CITROME, L. (2014). Quantifying clinical relevance. *Innov. Clin. Neurosci.* **11** 26–30.

Key words and phrases. Externally standardized residual, heterogeneity, replicability, statistical power, systematic review, meta-analysis.

OPEN SCIENCE COLLABORATION (2015). Estimating the reproducibility of psychological science. *Science* **349** aac4716. https://doi.org/10.1126/science.aac4716

NATURE COMMUNICATIONS (2022). Replication studies hold the key to generalization. *Nat. Commun.* **13** 7004. https://doi.org/10.1038/s41467-022-34748-x

DAVEY, J., TURNER, R. M., CLARKE, M. J. and HIGGINS, J. P. T. (2011). Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: A cross-sectional, descriptive analysis. *BMC Med. Res. Methodol.* **11** 160. https://doi.org/10.1186/1471-2288-11-160

DECHARTRES, A., ALTMAN, D. G., TRINQUART, L., BOUTRON, I. and RAVAUD, P. (2014). Association between analytic strategy and estimates of treatment outcomes in meta-analyses. *JAMA* **312** 623–630. https://doi.org/10.1001/jama.2014.8166

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

FRANCIS, G. (2012). The psychology of replication and replication in psychology. *Perspect. Psychol. Sci.* **7** 585–594. https://doi.org/10.1177/1745691612459520

GOODMAN, S. N. (2016). Aligning statistical and scientific reasoning: Misunderstanding and misuse of statistical significance impede science. *Science* **352** 1180–1181. MR3524721 https://doi.org/10.1126/science.aaf5406

GUMBEL, E. J. (1958). *Statistics of Extremes*. Columbia Univ. Press, New York. MR0096342

HEDGES, L. V. (2019). The statistics of replication. *Methodology* **15** 3–14. https://doi.org/10.1027/1614-2241/a000173

HEDGES, L. V. and SCHAUER, J. M. (2019a). More than one replication study is needed for unambiguous tests of replication. *J. Educ. Behav. Stat.* **44** 543–570. https://doi.org/10.3102/1076998619852953

HEDGES, L. V. and SCHAUER, J. M. (2019b). Statistical analyses for studying replication: Meta-analytic perspectives. *Psychol. Methods* **24** 557–570. https://doi.org/10.1037/met0000189

HELD, L. (2020). A new standard for the analysis and design of replication studies. *J. Roy. Statist. Soc. Ser. A* **183** 431–448. MR4052785 https://doi.org/10.1111/rssa.12493

HELD, L., MATTHEWS, R., OTT, M. and PAWEL, S. (2022). Reverse-Bayes methods for evidence assessment and research synthesis. *Res. Synth. Methods* **13** 295–314. https://doi.org/10.1002/jrsm.1538

HIGGINS, J. P. T., THOMAS, J., CHANDLER, J., CUMPSTON, M., LI, T., PAGE, M. J. and WELCH, V. A. (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley, New York. https://doi.org/10.1002/9781119536604

HIGGINS, J. P. T., THOMPSON, S. G., DEEKS, J. J. and ALTMAN, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ* **327** 557–560. https://doi.org/10.1136/bmj.327.7414.557

HIGGINS, J. P. T., THOMPSON, S. G. and SPIEGELHALTER, D. J. (2009). A re-evaluation of random-effects meta-analysis. *J. Roy. Statist. Soc. Ser. A* **172** 137–159. MR2655609 https://doi.org/10.1111/j.1467-985X.2008.00552.x

HOAGLIN, D. C. (2016). Misunderstandings about Q and 'Chochran's Q test' in meta-analysis. *Stat. Med.* **35** 485–495. MR3455577 https://doi.org/10.1002/sim.6632

HONG, C., SALANTI, G., MORTON, S. C., RILEY, R. D., CHU, H., KIMMEL, S. E. and CHEN, Y. (2020). Testing small study effects in multivariate meta-analysis. *Biometrics* **76** 1240–1250. MR4186839 https://doi.org/10.1111/biom.13342

HONG, H., CHU, H., ZHANG, J. and CARLIN, B. P. (2016). A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Res. Synth. Methods* **7** 6–22. https://doi.org/10.1002/jrsm.1153

HUTTON, B., SALANTI, G., CALDWELL, D. M., CHAIMANI, A., SCHMID, C. H., CAMERON, C., IOANNIDIS, J. P. A., STRAUS, S., THORLUND, K. et al. (2015). The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: Checklist and explanations. *Ann. Intern. Med.* **162** 777–784. https://doi.org/10.7326/M14-2385

IOANNIDIS, J. P. A. (2015). Failure to replicate: Sound the alarm. *Cerebrum* **2015**.

JACKSON, D. and WHITE, I. R. (2018). When should meta-analysis avoid making hidden normality assumptions? *Biom. J.* **60** 1040–1058. MR3876351 https://doi.org/10.1002/bimj.201800071

JALJULI, I., BENJAMINI, Y., SHENHAV, L., PANAGIOTOU, O. A. and HELLER, R. (2022). Quantifying replicability and consistency in systematic reviews. *Stat. Biopharm. Res.* **15** 372–385. https://doi.org/10.1080/19466315.2022.2050291

KARUNANANTHAN, S., MAXWELL, L. J., WELCH, V., PETKOVIC, J., PARDO PARDO, J., RADER, T., AVEY, M. T., BAPTISTE-NGOBI, J., BATISTA, R. et al. (2020). When and how to replicate systematic reviews. *Cochrane Database Syst. Rev.* **2** Art. No.: MR000052. https://doi.org/10.1002/14651858.MR000052

KLEIN, R. A., COOK, C. L., EBERSOLE, C. R., VITIELLO, C. A., NOSEK, B. A., CHARTIER, C. R., CHRISTOPHERSON, C. D., CLAY, S., COLLISSON, B. et al. (2019). Many Labs 4: Failure to replicate mortality salience effect with and without original author involvement. PsyArXiv. Available at https://psyarxiv.com/vef2c.

KLEIN, R. A., RATLIFF, K. A., VIANELLO, M., ADAMS JR, R. B., BAHNÍK, Š., BERNSTEIN, M. J., BOCIAN, K., BRANDT, M. J., BROOKS, B. et al. (2014). Investigating variation in replicability. *Soc. Psychol.* **45** 142–152. https://doi.org/10.1027/1864-9335/a000178

KULINSKAYA, E., DOLLINGER, M. B. and BJØRKESTØL, K. (2011). Testing for homogeneity in meta-analysis I. The one-parameter case: Standardized mean difference. *Biometrics* **67** 203–212. MR2898832 https://doi.org/10.1111/j.1541-0420.2010.01442.x

LAKENS, D., HILGARD, J. and STAAKS, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychol* **4** 24. https://doi.org/10.1186/s40359-016-0126-3

LANGAN, D., HIGGINS, J. P. T., JACKSON, D., BOWDEN, J., VERONIKI, A. A., KONTOPANTELIS, E., VIECHTBAUER, W. and SIMMONDS, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res. Synth. Methods* **10** 83–98. https://doi.org/10.1002/jrsm.1316

LIN, L. (2018). Bias caused by sampling error in meta-analysis with small sample sizes. *PLoS ONE* **13** e0204056. https://doi.org/10.1371/journal.pone.0204056

LIN, L. and CHU, H. (2018). Quantifying publication bias in meta-analysis. *Biometrics* **74** 785–794. MR3860700 https://doi.org/10.1111/biom.12815

LIN, L., CHU, H. and HODGES, J. S. (2016). Sensitivity to excluding treatments in network meta-analysis. *Epidemiology* **27** 562–569. https://doi.org/10.1097/EDE.0000000000000482

LIN, L., CHU, H. and HODGES, J. S. (2017). Alternative measures of between-study heterogeneity in meta-analysis: Reducing the impact of outlying studies. *Biometrics* **73** 156–166. MR3632361 https://doi.org/10.1111/biom.12543

MATHUR, M. B. and VANDERWEELE, T. J. (2019). Challenges and suggestions for defining replication "success" when effects may be heterogeneous: Comment on Hedges and Schauer (2019). *Psychol. Methods* **24** 571–575. https://doi.org/10.1037/met0000223

MATHUR, M. B. and VANDERWEELE, T. J. (2020). New statistical metrics for multisite replication projects. *J. Roy. Statist. Soc. Ser. A* **183** 1145–1166. MR4114480 https://doi.org/10.1111/rssa.12572

MILLS, E. J., IOANNIDIS, J. P. A., THORLUND, K., SCHÜNEMANN, H. J., PUHAN, M. A. and GUYATT, G. H. (2012). How to use an article reporting a multiple treatment comparison meta-analysis. *JAMA* **308** 1246–1253. https://doi.org/10.1001/2012.jama.11228

MILLS, E. J., KANTERS, S., THORLUND, K., CHAIMANI, A., VERONIKI, A.-A. and IOANNIDIS, J. P. A. (2013). The effects of excluding treatments from network meta-analyses: Survey. *BMJ* **347** f5195. https://doi.org/10.1136/bmj.f5195

MØLLER, C. H., PENNINGA, L., WETTERSLEV, J., STEINBRÜCHEL, D. A. and GLUUD, C. (2012). Off-pump versus on-pump coronary artery bypass grafting for ischaemic heart disease. *Cochrane Database Syst. Rev.* **3** Art. No.: CD007224. https://doi.org/10.1002/14651858.CD007224.pub2

MORRIS, T. P., WHITE, I. R. and CROWTHER, M. J. (2019). Using simulation studies to evaluate statistical methods. *Stat. Med.* **38** 2074–2102. MR3937487 https://doi.org/10.1002/sim.8086

MURAD, M. H., ASI, N., ALSAWAS, M. and ALAHDAB, F. (2016). New evidence pyramid. *BMJ Evid.-Based Med.* **21** 125–127. https://doi.org/10.1136/ebmed-2016-110401

NATIONAL ACADEMY OF SCIENCES (2019). *Reproducibility and Replicability in Science*. National Academies Press, Washington, DC. https://doi.org/10.17226/25303

NOORTGATE, W. V. D. and ONGHENA, P. (2005). Parametric and nonparametric bootstrap methods for meta-analysis. *Behav. Res. Methods* **37** 11–22. https://doi.org/10.3758/bf03206394

NOSEK, B. A. and ERRINGTON, T. M. (2020). What is replication? *PLoS Biol.* **18** e3000691. https://doi.org/10.1371/journal.pbio.3000691

NOSEK, B. A., HARDWICKE, T. E., MOSHONTZ, H., ALLARD, A., CORKER, K. S., DREBER, A., FIDLER, F., HILGARD, J., KLINE STRUHL, M. et al. (2022). Replicability, robustness, and reproducibility in psychological science. *Annu. Rev. Psychol.* **73** 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

PATIL, P., PENG, R. D. and LEEK, J. T. (2016). What should researchers expect when they replicate studies? A statistical view of replicability in psychological science. *Perspect. Psychol. Sci.* **11** 539–544. https://doi.org/10.1177/1745691616646366

SALANTI, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: Many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res. Synth. Methods* **3** 80–97. https://doi.org/10.1002/jrsm.1037

SCHAUER, J. M. and HEDGES, L. V. (2020). Assessing heterogeneity and power in replications of psychological experiments. *Psychol. Bull.* **146** 701–719. https://doi.org/10.1037/bul0000232

SCHMIDT, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Rev. Gen. Psychol.* **13** 90–100. https://doi.org/10.1037/a0015108

SIDDAWAY, A. P., WOOD, A. M. and HEDGES, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annu. Rev. Psychol.* **70** 747–770. https://doi.org/10.1146/annurev-psych-010418-102803

SPIEGELHALTER, D. J., ABRAMS, K. R. and MYLES, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, Hoboken, NJ.

STANLEY, T. D., CARTER, E. C. and DOUCOULIAGOS, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychol. Bull.* **144** 1325–1346. https://doi.org/10.1037/bul0000169

TAKKOUCHE, B., CADARSO-SUÁREZ, C. and SPIEGELMAN, D. (1999). Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis. *Amer. J. Epidemiol.* **150** 206–215. https://doi.org/10.1093/oxfordjournals.aje.a009981

TUGWELL, P., WELCH, V. A., KARUNANANTHAN, S., MAXWELL, L. J., AKL, E. A., AVEY, M. T., BHUTTA, Z. A., BROUWERS, M. C., CLARK, J. P. et al. (2020). When to replicate systematic reviews of interventions: Consensus checklist. *BMJ* **370** m2864. https://doi.org/10.1136/bmj.m2864

USEEM, J., BRENNAN, A., LAVALLEY, M., VICKERY, M., AMELI, O., REINEN, N. and GILL, C. J. (2015). Systematic differences between cochrane and non-cochrane meta-analyses on the same topic: A matched pair analysis. *PLoS ONE* **10** e0144980. https://doi.org/10.1371/journal.pone.0144980

VERONIKI, A. A., JACKSON, D., VIECHTBAUER, W., BENDER, R., BOWDEN, J., KNAPP, G., KUSS, O., HIGGINS, J. P. T., LANGAN, D. et al. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res. Synth. Methods* **7** 55–79. https://doi.org/10.1002/jrsm.1164

VIECHTBAUER, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36** 3. https://doi.org/10.18637/jss.v036.i03

VON HIPPEL, P. T. (2015). The heterogeneity statistic I(2) can be biased in small meta-analyses. *BMC Med. Res. Methodol.* **15** 35. https://doi.org/10.1186/s12874-015-0024-z

WANG, J. and OWEN, A. B. (2019). Admissibility in partial conjunction testing. *J. Amer. Statist. Assoc.* **114** 158–168. MR3941245 https://doi.org/10.1080/01621459.2017.1385465

WHITEHEAD, A. and WHITEHEAD, J. (1991). A general parametric approach to the meta-analysis of randomized clinical trials. *Stat. Med.* **10** 1665–1677. https://doi.org/10.1002/sim.4780101105

XIAO, M., CHU, H., HODGES, J. S. and LIN, L. (2024a). Supplement to “Quantifying replicability of multiple studies in a meta-analysis.” https://doi.org/10.1214/23-AOAS1806SUPPA

XIAO, M., CHU, H., HODGES, J. S. and LIN, L. (2024b). R package for “Quantifying replicability of multiple studies in a meta-analysis.” https://doi.org/10.1214/23-AOAS1806SUPPB

XING, A., CHU, H. and LIN, L. (2020). Fragility index of network meta-analysis with application to smoking cessation data. *J. Clin. Epidemiol.* **127** 29–39. https://doi.org/10.1016/j.jclinepi.2020.07.003

ZHANG, J., CARLIN, B. P., NEATON, J. D., SOON, G. G., NIE, L., KANE, R., VIRNIG, B. A. and CHU, H. (2014). Network meta-analysis of randomized clinical trials: Reporting the proper summaries. *Clin. Trials* **11** 246–262. https://doi.org/10.1177/1740774513498322

ZHANG, J., YUAN, Y. and CHU, H. (2016). The impact of excluding trials from network meta-analyses—An empirical study. *PLoS ONE* **11** e0165889. https://doi.org/10.1371/journal.pone.0165889

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 683–703
https://doi.org/10.1214/23-AOAS1807

MODELING THE VISIBILITY DISTRIBUTION FOR RESPONDENT-DRIVEN SAMPLING WITH APPLICATION TO POPULATION SIZE ESTIMATION

By Katherine R. McLaughlin[1,a], Lisa G. Johnston[2,b], Xhevat Jakupi[3,c], Dafina Gexha-Bunjaku[3,d], Edona Deva[4,e] and Mark S. Handcock[4,f]

[1]*Department of Statistics, Oregon State University,* [a]*katherine.mclaughlin@oregonstate.edu*
[2]*LGJ Consultants, Inc.,* [b]*lsjohnston.global@gmail.com*
[3]*National Institute of Public Health of Kosovo,* [c]*xhevat.r.jakupi@rks-gov.net,* [d]*dafinagexha@gmail.com*
[4]*Community Development Fund,* [e]*edona.deva@kcdf.org*
[5]*Department of Statistics and Data Science, University of California, Los Angeles,* [f]*handcock@stat.ucla.edu*

Respondent-driven sampling (RDS) is used throughout the world to estimate prevalence and population size for hidden populations. Although RDS is an effective method for enrolling people from key populations in studies, it relies on a partially unknown sampling mechanism, and thus each individual's inclusion probability is unknown. Current estimators for population prevalence, population size, and other outcomes rely on a participant's network size (degree) to approximate their inclusion probability in the sample from the networked population. However, in most RDS studies, a participant's network size is attained via a self-report and is subject to many types of misreporting and bias. Because design-based inclusion probabilities cannot be exactly computed, we instead use the term visibility to describe how likely a person is to be selected to participate in the study. The commonly used successive sampling population size estimation (SS-PSE) framework to estimate population sizes from RDS data relies on self-reported network sizes in the model for the sampling mechanism. We propose an enhancement of the SS-PSE framework that adds a measurement error model for visibility used in place of the self-reported network size and a model for the number of recruits an individual can enroll. Inferred visibilities are a way to smooth the degree distribution and bring in outliers as well as a mechanism to deal with missing and invalid network sizes. We demonstrate the performance of visibility SS-PSE on three populations from Kosovo sampled in 2014 using RDS. We also discuss how the visibility modeling framework could be extended to prevalence estimation.

REFERENCES

Andreatta, G. and Kaufman, G. M. (1986). Estimation of finite population properties when sampling is without replacement and proportional to magnitude. *J. Amer. Statist. Assoc.* **81** 657–666. MR0860497

Bell, D. C., Belli-McQueen, B. and Haider, A. (2007). Partner naming and forgetting: Recall of network members. *Soc. Netw.* **29** 279–299. https://doi.org/10.1016/j.socnet.2006.12.004

Bengtsson, L. and Thorson, A. (2010). Global HIV surveillance among MSM: Is risk behavior seriously underestimated? *AIDS* **24** 2301–2303. https://doi.org/10.1097/QAD.0b013e32833d207d

Bickel, P. J., Nair, V. N. and Wang, P. C. C. (1992). Nonparametric inference under biased sampling from a finite population. *Ann. Statist.* **20** 853–878. MR1165596 https://doi.org/10.1214/aos/1176348660

Brewer, D. D. (2000). Forgetting in the recall-based elicitation of personal and social networks. *Soc. Netw.* **22** 29–43. https://doi.org/10.1016/S0378-8733(99)00017-9

Conway, R. W. and Maxwell, W. L. (1962). A queuing model with state dependent service rates. *Int. J. Ind. Eng.* **12** 132–136.

Fellows, I. E. (2019). Respondent-driven sampling and the homophily configuration graph. *Stat. Med.* **38** 131–150. MR3887272 https://doi.org/10.1002/sim.7973

Key words and phrases. Heaped data, hidden population, measurement error model, model-based survey sampling, network sampling.

FELLOWS, I. E. (2022a). Estimating population size from a privatized network sample. *J. Surv. Stat. Methodol.* **10** 1346–1369. https://doi.org/10.1093/jssam/smac010

FELLOWS, I. E. (2022b). On the robustness of respondent-driven sampling estimators to measurement error. *J. Surv. Stat. Methodol.* **10** 377–396. https://doi.org/10.1093/jssam/smab056

FENTON, K. A., JOHNSON, A. M., MCMANUS, S. and ERENS, B. (2001). Measuring sexual behaviour: Methodological challenges in survey research. *Sex. Transm. Infect.* **77** 84–92. https://doi.org/10.1136/sti.77.2.84

FISHER, R. J. (1993). Social desirability bias and the validity of indirect questioning. *J. Consum. Res.* **20** 303. https://doi.org/10.1086/209351

GEXHA BUNJAKU, D., DEVA, E., GASHI, L., KAÇANIKU-GUNGA, P., COMINS, C. A. and EMMANUEL, F. (2019). Programmatic mapping to estimate size, distribution, and dynamics of key populations in Kosovo. *JMIR Public Health Surveill.* **5** e11194. https://doi.org/10.2196/11194

GILE, K. J. (2011). Improved inference for respondent-driven sampling data with application to HIV prevalence estimation. *J. Amer. Statist. Assoc.* **106** 135–146. MR2816708 https://doi.org/10.1198/jasa.2011.ap09475

GILE, K. J., BEAUDRY, I. S., HANDCOCK, M. S. and OTT, M. Q. (2018). Methods for inference from respondent-driven sampling data. *Annu. Rev. Stat. Appl.* **5** 65–96. MR3774740 https://doi.org/10.1146/annurev-statistics-031017-100704

GILE, K. J. and HANDCOCK, M. S. (2010). Respondent-driven sampling: An assessment of current methodology. *Sociol. Method.* **40** 285–327. https://doi.org/10.1111/j.1467-9531.2010.01223.x

GILE, K. J., JOHNSTON, L. G. and SALGANIK, M. J. (2015). Diagnostics for respondent-driven sampling. *J. Roy. Statist. Soc. Ser. A* **178** 241–269. MR3291770 https://doi.org/10.1111/rssa.12059

GROVES, R. M., FOWLER, F. J. JR., COUPER, M. P., LEPKOWSKI, J. M., SINGER, E. and TOURANGEAU, R. (2009). *Survey Methodology*, 2nd ed. *Wiley Series in Survey Methodology*. Wiley, Hoboken, NJ. MR3241261

HANDCOCK, M. S. and GILE, K. J. (2010). Modeling networks from sampled data. *Ann. Appl. Stat.* **272** 383–426.

HANDCOCK, M. S., GILE, K. J., KIM, B. J. and MCLAUGHLIN, K. R. (2022). sspse: Estimating hidden population size using respondent driven sampling data, Los Angeles, CA. R package version 1.1.0.

HANDCOCK, M. S., GILE, K. J. and MAR, C. M. (2014). Estimating hidden population size using respondent-driven sampling data. *Electron. J. Stat.* **8** 1491–1521. MR3263129 https://doi.org/10.1214/14-EJS923

HANDCOCK, M. S., GILE, K. J. and MAR, C. M. (2015). Estimating the size of populations at high risk for HIV using respondent-driven sampling data. *Biometrics* **71** 258–266. MR3335370 https://doi.org/10.1111/biom.12255

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

JOHNSTON, L. G. (2013). *Introduction to Respondent-Driven Sampling*. World Health Organization, Geneva, Switzerland.

JOHNSTON, L. G., MCLAUGHLIN, K. R., EL RHILANI, H., LATIFI, A., TOUFIK, A., BENNANI, A., ALAMI, K., ELOMARI, B. and HANDCOCK, M. S. (2015). Estimating the size of hidden populations using respondent-driven sampling data: Case examples from Morocco. *Epidemiology* **26** 846–852.

JOHNSTON, L. G., MCLAUGHLIN, K. R., GIOS, L., CORDIOLI, M., STANEKOVÁ, D. V., BLONDEEL, K., TOSKIN, I., MIRANDOLA, M. and FOR THE SIALON II NETWORK* (2021). Populations size estimations using SS-PSE among MSM in four European cities: How many MSM are living with HIV? *Eur. J. Public Health* **31** 1129–1136. https://doi.org/10.1093/eurpub/ckab148

JOHNSTON, L. G., MCLAUGHLIN, K. R., ROUHANI, S. A. and BARTELS, S. A. (2017). Measuring a hidden population: A novel technique to estimate the population size of women with sexual violence-related pregnancies in South Kivu Province, Democratic Republic of Congo. *J. Epidemiol. Glob. Health* **7** 45–53. https://doi.org/10.1016/j.jegh.2016.08.003

JOHNSTON, L. G., WHITEHEAD, S., SIMIC-LAWSON, M. and KENDALL, C. (2010). Formative research to optimize respondent-driven sampling surveys among hard-to-reach populations in HIV behavioral and biological surveillance: Lessons learned from four case studies. *AIDS Care* **22** 784–792. https://doi.org/10.1080/09540120903373557

KHAN, B., LEE, H.-W., FELLOWS, I. and DOMBROWSKI, K. (2018). One-step estimation of networked population size: Respondent-driven capture-recapture with anonymity. *PLoS ONE* **13** 1–39. https://doi.org/10.1371/journal.pone.0195959

KIM, B. J. and HANDCOCK, M. S. (2019). Population size estimation using multiple respondent-driven sampling surveys. *J. Surv. Stat. Methodol.* **9** 94–120. https://doi.org/10.1093/jssam/smz055

KOKONENDJI, C. C. (2014). Over- and underdispersion models. In *Methods and Applications of Statistics in Clinical Trials. Vol.* 2. *Wiley Ser. Methods Appl. Statist.* 506–526. Wiley, Hoboken, NJ. MR3287469 https://doi.org/10.1002/9781118596333.ch30

KOSOVO HIV INTEGRATED BEHAVIORAL BIOLOGICAL SURVEILLANCE SURVEY REFERENCE GROUP (2014). HIV integrated behavioral and biological surveillance surveys—Kosovo. Technical report.

LEE, S., SUZER-GURTEKIN, T., WAGNER, J. and VALLIANT, R. (2017). Total survey error and respondent driven sampling: Focus on nonresponse and measurement errors in the recruitment process and the network size reports and implications for inferences. *J. Off. Stat.* **33** 335–366. https://doi.org/doi:10.1515/jos-2017-0017

LU, X. (2013). Linked ego networks: Improving estimate reliability and validity with respondent-driven sampling. *Soc. Netw.* **35** 669–685. https://doi.org/10.1016/j.socnet.2013.10.001

MCLAUGHLIN, K. R., JOHNSTON, L. G., GAMBLE, L. J., GRIGORYAN, T., PAPOYAN, A. and GRIGORYAN, S. (2019). Population size estimations among hidden populations using respondent-driven sampling surveys: Case studies from Armenia. *JMIR Public Health Surveill.* **5** e12034. https://doi.org/10.2196/12034

MCLAUGHLIN, K. R., JOHNSTON, L. G., JAKUPI, X., GEXHA-BUNJAKU, D., DEVA, E. and HANDCOCK, M. S. (2024). Supplement to "Modeling the visibility distribution for respondent-driven sampling with application to population size estimation." https://doi.org/10.1214/23-AOAS1807SUPPA, https://doi.org/10.1214/23-AOAS1807SUPPB

MENG, X.-L. (1994). Posterior predictive p-values. *Ann. Statist.* **22** 1142–1160. MR1311969 https://doi.org/10.1214/aos/1176325622

MILLS, H. L., JOHNSON, S., HICKMAN, M., JONES, N. S. and COLIJN, C. (2014). Errors in reported degrees and respondent driven sampling: Implications for bias. *Drug Alcohol Depend.* **142** 120–126. https://doi.org/10.1016/j.drugalcdep.2014.06.015

MURTHY, M. N. (1957). Ordered and unordered estimators in sampling without replacement. *Sankhyā* **18** 379–390. MR0094869

NAIR, V. N. and WANG, P. C. C. (1989). Maximum likelihood estimation under a successive sampling discovery model. *Technometrics* **31** 423–436. MR1041563 https://doi.org/10.2307/1269993

PLUMMER, M., BEST, N., COWLES, K. and VINES, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6** 7–11.

R CORE TEAM (2020). *R*: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RAJ, D. (1956). Some estimators in sampling with varying probabilities without replacement. *J. Amer. Statist. Assoc.* **51** 269–284. MR0082236

RIDOUT, M. S. and BESBEAS, P. (2004). An empirical model for underdispersed count data. *Stat. Model.* **4** 77–89. MR2037815 https://doi.org/10.1191/1471082X04st064oa

RUBIN, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.* **12** 1151–1172. MR0760681 https://doi.org/10.1214/aos/1176346785

SALGANIK, M. J. and HECKATHORN, D. D. (2004). Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol. Method.* **34** 193–240. https://doi.org/10.1111/j.0081-1750.2004.00152.x

TOMAS, A. and GILE, K. J. (2011). The effect of differential recruitment, non-response and non-recruitment on estimators for respondent-driven sampling. *Electron. J. Stat.* **5** 899–934. MR2831520 https://doi.org/10.1214/11-EJS630

VOLZ, E. and HECKATHORN, D. D. (2008). Probability based estimation theory for respondent-driven sampling. *J. Off. Stat.* **24** 79–97.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 704–728
https://doi.org/10.1214/23-AOAS1808

A PARTIALLY FUNCTIONAL LINEAR REGRESSION FRAMEWORK FOR INTEGRATING GENETIC, IMAGING, AND CLINICAL DATA

BY TING LI[1,a], YANG YU[2,b], J. S. MARRON[2,c] AND HONGTU ZHU[3,d]

[1] *School of Statistics and Management, Shanghai University of Finance and Economics,* [a] *tingli@mail.shufe.edu.cn*
[2] *Department of Statistics and Operations Research, University of North Carolina at Chapel Hill,* [b] *yangyu.unc@gmail.com,* [c] *marron@unc.edu*
[3] *Department of Biostatistics, University of North Carolina at Chapel Hill,* [d] *htzhu@email.unc.edu*

This paper is motivated by the joint analysis of genetic, imaging, and clinical (GIC) data collected in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. We propose a partially functional linear regression (PFLR) framework to map high-dimensional GIC-related pathways for Alzheimer's disease (AD). We develop a joint model selection and estimation procedure by embedding imaging data in the reproducing kernel Hilbert space and imposing the ℓ_0 penalty for the coefficients of genetic variables. We apply the proposed method to the ADNI dataset to identify important features from tens of thousands of genetic polymorphisms (reduced from millions using a preprocessing step) and study the effects of a certain set of informative genetic variants and the baseline hippocampus surface on 13 future cognitive scores. We also explore the shared and distinct heritability patterns of these cognitive scores. Analysis results suggest that both the hippocampal and genetic data have heterogeneous effects on different scores, with the trend that the value of both hippocampi are negatively associated with the severity of cognition deficits. Polygenic effects are observed for all the thirteen cognitive scores. The well-known APOE4 genotype only explains a small part of the cognitive function. Shared genetic etiology exists; however, greater genetic heterogeneity exists within disease classifications after accounting for the baseline diagnosis status. These analyses are useful in further investigation of functional mechanisms for AD progression.

REFERENCES

BATTISTA, P., SALVATORE, C. and CASTIGLIONI, I. (2017). Optimizing neuropsychological assessments for cognitive, behavioral, and functional impairment classification: A machine learning study. *Behav. Neurol.* **2017** 1–19.

BERTRAM, L. and TANZI, R. E. (2012). The genetics of Alzheimer's disease. *Prog. Mol. Biol. Transl. Sci.* **107** 79–100.

BRAAK, H. and BRAAK, E. (1998). *Evolution of Neuronal Changes in the Course of Alzheimer's Disease*. Springer, Berlin.

CAI, T. T. and YUAN, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist.* **39** 2330–2355.

CAI, T. T. and YUAN, M. (2012). Minimax and adaptive prediction for functional linear regression. *J. Amer. Statist. Assoc.* **107** 1201–1216.

CARRASQUILLO, M. M., ZOU, F., PANKRATZ, V. S., WILCOX, S. L., MA, L., WALKER, L. P., YOUNKIN, S. G., YOUNKIN, C. S., YOUNKIN, L. H. et al. (2009). Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nat. Genet.* **41** 192–198.

CRUCIANI, F., ALTMANN, A., LORENZI, M., MENEGAZ, G. and GALAZZO, I. B. (2022). What PLS can still do for imaging genetics in Alzheimer's disease. In 2022 *IEEE-EMBS International Conference on Biomedical and Health Informatics* (*BHI*) 1–4. IEEE.

DE AQUINO, C. H. (2021). Methodological issues in randomized clinical trials for prodromal Alzheimer's and Parkinson's disease. *Front. Neurol.* **12** 694329.

Key words and phrases. Clinical, genetics, imaging, nonasymptotic error bounds, partially functional linear regression, sparsity.

de Flores, R., Joie, R. L. and Chételat, G. (2015). Structural imaging of hippocampal subfields in healthy aging and Alzheimer's disease. *Neuroscience* **309** 29–50.

Deary, I. J., Cox, S. R. and Hill, W. D. (2022). Genetic variation, brain, and intelligence differences. *Mol. Psychiatry* **27** 335–353.

Dukart, J., Sambataro, F. and Bertolino, A. (2016). Accurate prediction of conversion to Alzheimer's disease using imaging, genetic, and neuropsychological biomarkers. *J. Alzheimer's Dis.* **49** 1143–1159.

Duong, M. T., Das, S. R., Lyu, X., Xie, L., Richardson, H., Xie, S. X., Yushkevich, P. A., Wolk, D. A. and Nasrallah, I. M. (2022). Dissociation of tau pathology and neuronal hypometabolism within the ATN framework of Alzheimer's disease. *Nat. Commun.* **13** 1–15.

Elliott, L. T., Sharp, K., Alfaro-Almagro, F., Shi, S., Miller, K. L., Douaud, G., Marchini, J. and Smith, S. M. (2018). Genome-wide association studies of brain imaging phenotypes in UK biobank. *Nature* **562** 210–216.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360.

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.

Frisoni, G. B., Ganzola, R., Canu, E., Rüb, U., Pizzini, F. B., Alessandrini, F., Zoccatelli, G., Beltramello, A., Caltagirone, C. et al. (2008). Mapping local hippocampal changes in Alzheimer's disease and normal ageing with MRI at 3 Tesla. *Brain* **131** 3266–3276.

Ge, T., Nichols, T. E., Lee, P. H., Holmes, A. J., Roffman, J. L., Buckner, R. L., Sabuncu, M. R. and Smoller, J. W. (2015). Massively expedited genome-wide heritability analysis (MEGHA). *Proc. Natl. Acad. Sci. USA* **112** 2479–2484.

Grassi, M., Rouleaux, N., Caldirola, D., Loewenstein, D., Schruers, K., Perna, G., Dumontier, M. and Initiative, A. D. N. (2019). A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimer's disease using socio-demographic characteristics, clinical information, and neuropsychological measures. *Front. Neurol.* 756.

Guerreiro, R. and Bras, J. (2015). The age factor in Alzheimer's disease. *Gen. Med.* **7** 1–3.

Hobart, J., Cano, S., Posner, H., Selnes, O., Stern, Y., Thomas, R., Zajicek, J. and Initiative, A. D. N. (2013). Putting the Alzheimer's cognitive test to the test I: Traditional psychometric methods. *Alzheimer's Dement.* **9** S4–S9.

Huang, C., Thompson, P., Wang, Y., Yu, Y., Zhang, J., Kong, D., Colen, R. R., Knickmeyer, R. C., Zhu, H. et al. (2017). FGWAS: Functional genome wide association analysis. *NeuroImage* **159** 107–121.

Huang, J., Jiao, Y., Liu, Y. and Lu, X. (2018). A constructive approach to L_0 penalized regression. *J. Mach. Learn. Res.* **19** 403–439.

Jack Jr., C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J. et al. (2013). Tracking pathophysiological processes in Alzheimer's disease: An updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* **12** 207–216.

Jack Jr., C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C. and Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* **9** 119–128.

Knutson, K. A., Deng, Y. and Pan, W. (2020). Implicating causal brain imaging endophenotypes in Alzheimer's disease using multivariable IWAS and GWAS summary data. *NeuroImage* **223** 117347.

Kong, D., Xue, K., Yao, F. and Zhang, H. H. (2016). Partially functional linear regression in high dimensions. *Biometrika* **103** 147–159.

Li, S., Shi, F., Pu, F., Li, X., Jiang, T., Xie, S. and Wang, Y. (2007). Hippocampal shape analysis of Alzheimer disease based on machine learning methods. *Am. J. Neuroradiol.* **28** 1339–1345.

Li, T., Yu, Y., Marron, J. and Zhu, H. (2024). Supplement to "A partially functional linear regression framework for integrating genetic, imaging, and clinical data." https://doi.org/10.1214/23-AOAS1808SUPPA, https://doi.org/10.1214/23-AOAS1808SUPPB

Li, T. and Zhu, Z. (2020). Inference for generalized partial functional linear regression. *Statist. Sinica* **30** 1379–1397.

Lin, D., Cao, H., Calhoun, V. D. and Wang, Y. P. (2014). Sparse models for correlative and integrative analysis of imaging and genetic data. *J. Neurosci. Methods* **237** 69–78.

Lo, M.-T., Kauppi, K., Fan, C.-C., Sanyal, N., Reas, E. T., Sundar, V., Lee, W.-C., Desikan, R. S., McEvoy, L. K. et al. (2019). Identification of genetic heterogeneity of Alzheimer's disease across age. *Neurobiol. Aging* **84** 243.e1–243.e9.

Ma, H., Li, T., Zhu, H. and Zhu, Z. (2019). Quantile regression for functional partially linear model in ultra-high dimensions. *Comput. Statist. Data Anal.* **129** 135–147.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack Jr., C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J. et al. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* **7** 263–269.

Morrison, C., Dadar, M., Shafiee, N., Villeneuve, S., Collins, D. L., Initiative, A. D. N. et al. (2022). Regional brain atrophy and cognitive decline depend on definition of subjective cognitive decline. *NeuroImage Clin.* **33** 102923.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack Jr., C. R., Jagust, W., Trojanowski, J. Q., Toga, A. W. and Beckett, L. (2005). Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's Dement.* **1** 55–66.

Nathoo, F. S., Kong, L. and Zhu, H. (2019). A review of statistical methods in imaging genetics. *Canad. J. Statist.* **47** 108–131.

Ossenkoppele, R., Leuzy, A., Cho, H., Sudre, C. H., Strandberg, O., Smith, R., Palmqvist, S., Mattsson-Carlgren, N., Olsson, T. et al. (2021). The impact of demographic, clinical, genetic, and imaging variables on tau PET status. *Eur. J. Nucl. Med. Mol. Imaging* **48** 2245–2258.

Pedraza, O., Bowers, D. and Gilmore, R. (2004). Asymmetry of the hippocampus and amygdala in MRI volumetric measurements of normal adults. *J. Int. Neuropsychol. Soc.* **10** 664–678.

Peng, G.-P., Feng, Z., He, F.-P., Chen, Z.-Q., Liu, X.-Y., Liu, P. and Luo, B.-Y. (2015). Correlation of hippocampal volume and cognitive performances in patients with either mild cognitive impairment or Alzheimer's disease. *CNS Neuroscience Ther.* **21** 15–22.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38** 904–909.

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

Selkoe, D. J. and Hardy, J. (2016). The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol. Med.* **8** 595–608.

Shen, L. and Thompson, P. M. (2020). Brain imaging genomics: Integrated analysis and machine learning. *Proc. IEEE Inst. Electr. Electron. Eng.* **108** 125–162.

Smith, S. M., Douaud, G., Chen, W., Hanayik, T., Alfaro-Almagro, F., Sharp, K. and Elliott, L. T. (2021). An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat. Neurosci.* **24** 737–745.

Thompson, P. M., Jahanshad, N., Ching, C. R., Salminen, L. E., Thomopoulos, S. I., Bright, J., Baune, B. T., Bertolín, S., Bralten, J. et al. (2020). ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl. Psychiatry* **10** 1–28.

Veitch, D. P., Weiner, M. W., Aisen, P. S., Beckett, L. A., DeCarli, C., Green, R. C., Harvey, D., Jack Jr., C. R., Jagust, W. et al. (2021). Using the Alzheimer's disease neuroimaging initiative to improve early detection, diagnosis, and treatment of Alzheimer's disease. *Alzheimer's Dement.* **18** 824–857.

Vermunt, L., Sikkes, S. A., Van Den Hout, A., Handels, R., Bos, I., Van Der Flier, W. M., Kern, S., Ousset, P.-J., Maruff, P. et al. (2019). Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype. *Alzheimer's Dement.* **15** 888–898.

Vina, J. and Lloret, A. (2010). Why women have more Alzheimer's disease than men: Gender and mitochondrial toxicity of amyloid-β peptide. *J. Alzheimer's Dis.* **20** S527–S533.

Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.

Wang, L., Kim, Y. and Li, R. (2013). Calibrating nonconvex penalized regression in ultra-high dimension. *Ann. Statist.* **41** 2505–2536.

Wilson, R. S., Barral, S., Lee, J. H., Leurgans, S. E., Foroud, T. M., Sweet, R. A., Graff-Radford, N., Bird, T. D., Mayeux, R. et al. (2011). Heritability of different forms of memory in the Late Onset Alzheimer's Disease Family Study. *J. Alzheimer's Dis.* **23** 249–255.

Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J. and Visscher, P. M. (2018). Common disease is more complex than implied by the core gene omnigenic model. *Cell* **173** 1573–1580.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G. et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42** 565–569.

Yao, F., Sue-Chee, S. and Wang, F. (2017). Regularized partially functional quantile regression. *J. Multivariate Anal.* **156** 39–56.

Yu, D., Wang, L., Kong, D. and Zhu, H. (2022). Mapping the genetic-imaging-clinical pathway with applications to Alzheimer's disease. *J. Amer. Statist. Assoc.* **117** 1656–1668.

YUAN, M. and CAI, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Statist.* **38** 3412–3444.

ZHAO, B., LI, T., SMITH, S. M., XIONG, D., WANG, X., YANG, Y., LUO, T., ZHU, Z., SHAN, Y. et al. (2022). Common variants contribute to intrinsic human brain functional networks. *Nat. Genet.* **54** 508–517.

ZHAO, B., LI, T., YANG, Y., WANG, X., LUO, T., SHAN, Y., ZHU, Z., XIONG, D., HAUBERG, M. E. et al. (2021). Common genetic variation influencing human white matter microstructure. *Science* **372** eabf3736.

ZHAO, B., LUO, T., LI, T., LI, Y., ZHANG, J., SHAN, Y., WANG, X., YANG, L., ZHOU, F. et al. (2019). Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat. Genet.* **51** 1637–1644.

ZHAO, B. and ZHU, H. (2022). On genetic correlation estimation with summary statistics from genome-wide association studies. *J. Amer. Statist. Assoc.* **117** 1–11.

ZHU, H., LI, T. and ZHAO, B. (2023). Statistical learning methods for neuroimaging data analysis with applications. *Annu. Rev. Biomed. Data Sci.* **6** 73–104.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 729–748
https://doi.org/10.1214/23-AOAS1809

COMPOSITE SCORES FOR TRANSPLANT CENTER EVALUATION: A NEW INDIVIDUALIZED EMPIRICAL NULL METHOD

BY NICHOLAS HARTMAN[1,a], JOSEPH M. MESSANA[2,e], JIAN KANG[1,b], ABHIJIT S. NAIK[2,f], TEMPIE H. SHEARON[1,c] AND KEVIN HE[1,d]

[1]*Department of Biostatistics, University of Michigan,* [a]*nhar@umich.edu,* [b]*jiankang@umich.edu,* [c]*tshearon@med.umich.edu,* [d]*kevinhe@umich.edu*

[2]*Division of Nephrology, University of Michigan,* [e]*jmessana@med.umich.edu,* [f]*abhinaik@med.umich.edu*

Risk-adjusted quality measures are used to evaluate healthcare providers with respect to national norms while controlling for factors beyond their control. Existing healthcare provider profiling approaches typically assume that the between-provider variation in these measures is entirely due to meaningful differences in quality of care. However, in practice, much of the between-provider variation will be due to trivial fluctuations in healthcare quality or unobservable confounding risk factors. If these additional sources of variation are not accounted for, conventional methods will disproportionately identify larger providers as outliers, even though their departures from the national norms may not be "extreme" or clinically meaningful. Motivated by efforts to evaluate the quality of care provided by transplant centers, we develop a composite evaluation score based on a novel individualized empirical null method, which robustly accounts for overdispersion due to unobserved risk factors, models the marginal variance of standardized scores as a function of the effective sample size, and only requires the use of publicly-available center-level statistics. The evaluations of United States kidney transplant centers based on the proposed composite score are substantially different from those based on conventional methods. Simulations show that the proposed empirical null approach more accurately classifies centers in terms of quality of care, compared to existing methods.

REFERENCES

ANGRIST, J. D., IMBENS, G. W. and RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* **91** 444–455.

ASH, A. S., FIENBERG, S. F., LOUIS, T. A., NORMAND, S. T., STUKEL, T. A. and UTTS, J. (2012). Statistical issues in assessing hospital performance. *Commissioned by the Committee of Presidents of Statistical Societies for the Centers for Medicare and Medicaid Services (CMS)* [online]. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf.

CARLIN, B. P. and LOUIS, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed. New York: Chapman & Hall/CRC.

CHEN, Y., RHEE, C., SENTURK, D., KURUM, E., CAMPOS, L., LI, Y., KALANTAR-ZADEH, K. and NGUYEN, D. (2019). Association of US dialysis facility staffing with profiling of hospital-wide 30-day unplanned readmission. *Kidney Dis.* **5** 153–162.

EFRON, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *J. Amer. Statist. Assoc.* **99** 96–104. MR2054289 https://doi.org/10.1198/016214504000000089

EFRON, B. (2007). Size, power and false discovery rates. *Ann. Statist.* **35** 1351–1377. MR2351089 https://doi.org/10.1214/009053606000001460

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

ESTES, J. P., CHEN, Y., ŞENTÜRK, D., RHEE, C. M., KÜRÜM, E., YOU, A. S., STREJA, E., KALANTAR-ZADEH, K. and NGUYEN, D. V. (2020). Profiling dialysis facilities for adverse recurrent events. *Stat. Med.* **39** 1374–1389. MR4098496 https://doi.org/10.1002/sim.8482

Key words and phrases. Empirical null, end-stage renal disease, provider profiling, unmeasured confounders.

ESTES, J. P., NGUYEN, D. V., CHEN, Y., DALRYMPLE, L. S., RHEE, C. M., KALANTAR-ZADEH, K. and ŞENTÜRK, D. (2018). Time-dynamic profiling with application to hospital readmission among patients on dialysis. *Biometrics* **74** 1383–1394. MR3908156 https://doi.org/10.1111/biom.12908

HART, A., SMITH, J. M., SKEANS, M. A., GUSTAFSON, S. K., WILK, A. R., CASTRO, S., FOUTZ, J., WAINRIGHT, J. L., SNYDER, J. J. et al. (2020). OPTN/SRTR 2018 annual data report: Kidney. *Am. J. Transplant.* **20** 20–130.

HARTMAN, N., MESSANA, J. M., KANG, J., NAIK, A. S., SHEARON, T. H. and HE, K. (2024). Supplement to "Composite scores for transplant center evaluation: A new individualized empirical null method." https://doi.org/10.1214/23-AOAS1809SUPPA, https://doi.org/10.1214/23-AOAS1809SUPPB

HE, K., DAHLERUS, C., XIA, L., LI, Y. and KALBFLEISCH, J. D. (2020). The profile inter-unit reliability. *Biometrics* **76** 654–663. MR4125285 https://doi.org/10.1111/biom.13167

HE, K., KALBFLEISCH, J. D., LI, Y. and LI, Y. (2013). Evaluating hospital readmission rates in dialysis facilities; adjusting for hospital effects. *Lifetime Data Anal.* **19** 490–512. MR3119994 https://doi.org/10.1007/s10985-013-9264-6

HUBER, P. J. (1981). *Robust Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0606374

JAY, C. and SCHOLD, J. D. (2017). Measuring transplant center performance: The goals are not controversial but the methods and consequences can be. *Curr. Transplant. Rep.* **4** 52–58. https://doi.org/10.1007/s40472-017-0138-9

JONES, H. E. and SPIEGELHALTER, D. J. (2011). The identification of "unusual" health-care providers from a hierarchical model. *Amer. Statist.* **65** 154–163. MR2848190 https://doi.org/10.1198/tast.2011.10190

KALBFLEISCH, J. D., HE, K., XIA, L. and LI, Y. M. (2018). Does the Inter-Unit Reliability (IUR) measure reliability? *Health Serv. Outcomes Res. Methodol.* **18** 215–225.

KALBFLEISCH, J. D. and WOLFE, R. A. (2013). On monitoring outcomes of medical providers. *Stat. Biosci.* **5** 286–302.

LIU, D., SCHAUBEL, D. E. and KALBFLEISCH, J. D. (2012). Computationally efficient marginal models for clustered recurrent event data. *Biometrics* **68** 637–647. MR2959631 https://doi.org/10.1111/j.1541-0420.2011.01676.x

MARRERO, W. J., NAIK, A. S., FRIEDEWALD, J. J., XU, Y., HUTTON, D. W., LAVIERI, M. S. and PARIKH, N. D. (2017). Predictors of deceased donor kidney discard in the United States. *Transplantation* **101** 1690–1697.

NELDER, J. A. and MEAD, R. (1965). A simplex method for function minimization. *Comput. J.* **7** 308–313. MR3363409 https://doi.org/10.1093/comjnl/7.4.308

O'CONNOR, K. and GLAZIER, A. (2021). OPO performance improvement and increasing organ transplantation: Metrics are necessary but not sufficient. *Am. J. Transplant.* **21** 2325–2326. https://doi.org/10.1111/ajt.16545

PROUDLOVE, N. C., GOFF, M., WALSHE, K. and BOADEN, R. (2019). The signal in the noise: Robust detection of performance 'outliers' in health services. *J. Oper. Res. Soc.* **70** 1102–1114.

SCIENTIFIC REGISTRY OF TRANSPLANT RECIPIENTS (2021a). SRTR Risk Adjustment Model Documentation: Waiting List Models. Available at https://www.srtr.org/tools/waiting-list/.

SCIENTIFIC REGISTRY OF TRANSPLANT RECIPIENTS (2021b). Technical Methods for the Program-Specific Reports. Available at https://www.srtr.org/about-the-data/technical-methods-for-the-program-specific-reports/.

SHWARTZ, M., RESTUCCIA, J. D. and ROSEN, A. K. (2015). Composite measures of health care provider performance: A description of approaches. *Milbank Q.* **93** 788–825. https://doi.org/10.1111/1468-0009.12165

SPIEGELHALTER, D., SHERLAW-JOHNSON, C., BARDSLEY, M., BLUNT, I., WOOD, C. and GRIGG, O. (2012). Statistical methods for healthcare regulation: Rating, screening and surveillance. *J. Roy. Statist. Soc. Ser. A* **175** 1–47. MR2873790 https://doi.org/10.1111/j.1467-985X.2011.01010.x

SPIEGELHALTER, D. J. (2005). Funnel plots for comparing institutional performance. *Stat. Med.* **24** 1185–1202. MR2134573 https://doi.org/10.1002/sim.1970

UNITED STATES RENAL DATA SYSTEM (2021). *2021 USRDS Annual Data Report: Epidemiology of Kidney Disease in the United States*. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD.

WOLFE, R. A., ASHBY, V. B., MILFORD, E. L., OJO, A. O., ETTENGER, R. E., AGODOA, L. Y. C., HELD, P. J. and PORT, F. K. (1999). Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *N. Engl. J. Med.* **341** 1725–1730.

XIA, L., HE, K., LI, Y. and KALBFLEISCH, J. (2022). Accounting for total variation and robustness in profiling health care providers. *Biostatistics* **23** 257–273. MR4366047 https://doi.org/10.1093/biostatistics/kxaa024

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 749–769
https://doi.org/10.1214/23-AOAS1810

A CONTINUOUS MULTIPLE HYPOTHESIS TESTING FRAMEWORK FOR OPTIMAL EXOPLANET DETECTION

BY NATHAN C. HARA[1,a], THIBAULT DE POYFERRÉ[2,c], JEAN-BAPTISTE DELISLE[1,b] AND MARC HOFFMANN[3,d]

[1]*Observatoire Astronomique de l'Université de Genève,* [a]*nathan.hara@unige.ch,* [b]*jean-baptiste.delisle@unige.ch*
[2]*Mathematical Science Research Institute, University of California, Berkeley,* [c]*thibault.de.poyferre@gmail.com*
[3]*University Paris–Dauphine, CEREMADE,* [d]*hoffmann@ceremade.dauphine.fr*

When searching for exoplanets, one wants to count how many planets orbit a given star, and to determine what their characteristics are. If the estimated planet characteristics are too far from those of a planet truly present, this should be considered as a false detection. This setting is a particular instance of a general one: aiming to retrieve parametric components in a dataset corrupted by nuisance signals, with a certain accuracy on their parameters. We exhibit a detection criterion minimizing false and missed detections, either as a function of their relative cost or when the expected number of false detections is bounded. If the components can be separated in a technical sense discussed in detail, the optimal detection criterion is a posterior probability obtained as a by-product of Bayesian evidence calculations. Optimality is guaranteed within a model, and we introduce model criticism methods to ensure that the criterion is robust to model errors. We show on two simulations emulating exoplanet searches that the optimal criterion can significantly outperform other criteria. Finally, we show that our framework offers solutions for the identification of components of mixture models and Bayesian false discovery rate control when hypotheses are not discrete.

REFERENCES

AHRER, E., QUELOZ, D., RAJPAUL, V. M., SÉGRANSAN, D., BOUCHY, F., HALL, R., HANDLEY, W., LOVIS, C., MAYOR, M. et al. (2021). The HARPS search for southern extra-solar planets—XLV. Two Neptune mass planets orbiting HD 13808: A study of stellar activity modelling's impact on planet detection. *Mon. Not. R. Astron. Soc.* **503** 1248–1263. https://doi.org/10.1093/mnras/stab373

AIGRAIN, S., PONT, F. and ZUCKER, S. (2012). A simple method to estimate radial velocity variations due to stellar activity using photometry. *Mon. Not. R. Astron. Soc.* **419** 3147–3158. https://doi.org/10.1111/j.1365-2966.2011.19960.x

BALUEV, R. V. (2008). Assessing the statistical significance of periodogram peaks. *Mon. Not. R. Astron. Soc.* **385** 1279–1285. https://doi.org/10.1111/j.1365-2966.2008.12689.x

BALUEV, R. V. (2009). Accounting for velocity jitter in planet search surveys. *Mon. Not. R. Astron. Soc.* **393** 969–978. https://doi.org/10.1111/j.1365-2966.2008.14217.x

BALUEV, R. V. (2013). Detecting multiple periodicities in observational data with the multifrequency periodogram—I. Analytic assessment of the statistical significance. *Mon. Not. R. Astron. Soc.* **436** 807–818. https://doi.org/10.1093/mnras/stt1617

BALUEV, R. V. (2015). Keplerian periodogram for Doppler exoplanet detection: Optimized computation and analytic significance thresholds. *Mon. Not. R. Astron. Soc.* **446** 1478–1492. https://doi.org/10.1093/mnras/stu2191

BARBIERI, M. M. and BERGER, J. O. (2004). Optimal predictive model selection. *Ann. Statist.* **32** 870–897. MR2065192 https://doi.org/10.1214/009053604000000238

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

BOGDAN, M., GHOSH, J. K. and TOKDAR, S. T. (2008). A comparison of the Benjamini–Hochberg procedure with some Bayesian rules for multiple testing. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen. Inst. Math. Stat. (IMS) Collect.* **1** 211–230. IMS, Beachwood, OH. MR2462208 https://doi.org/10.1214/193940307000000158

Key words and phrases. Exoplanets, Bayesian, false discovery rate, decision theory, maximum utility.

BOX, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *J. Roy. Statist. Soc. Ser. A* **143** 383–430. With discussion. MR0603745 https://doi.org/10.2307/2982063

BREWER, B. J. and DONOVAN, C. P. (2015). Fast Bayesian inference for exoplanet discovery in radial velocity data. *Mon. Not. R. Astron. Soc.* **448** 3206–3214. https://doi.org/10.1093/mnras/stv199

CELEUX, G., FRÜHWIRTH-SCHNATTER, S. and ROBERT, C. P. (2018). Model selection for mixture models—perspectives and strategies. In *Handbook of Mixture Analysis*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods*. CRC Press, Boca Raton, FL. MR3889692

CELEUX, G., HURN, M. and ROBERT, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *J. Amer. Statist. Assoc.* **95** 957–970. MR1804450 https://doi.org/10.2307/2669477

COVER, T. M. and THOMAS, J. A. (1991). *Elements of Information Theory*. *Wiley Series in Telecommunications*. Wiley, New York. A Wiley-Interscience Publication. MR1122806 https://doi.org/10.1002/0471200611

CRASS, J., GAUDI, B. S., LEIFER, S., BEICHMAN, C., BENDER, C., BLACKWOOD, G., BURT, J. A., CALLAS, J. L., CEGLA, H. M. et al. (2021). Extreme precision radial velocity working group final report. arXiv e-prints. Available at arXiv:2107.14291.

CRETIGNIER, M., DUMUSQUE, X., HARA, N. C. and PEPE, F. (2021). YARARA: Significant improvement in RV precision through post-processing of spectral time series. *A & A* **653** A43. https://doi.org/10.1051/0004-6361/202140986

CUMMING, A. (2004). Detectability of extrasolar planets in radial velocity surveys. *Mon. Not. R. Astron. Soc.* **354** 1165–1176. https://doi.org/10.1111/j.1365-2966.2004.08275.x

DAWID, A. P. (1982). The well-calibrated Bayesian. *J. Amer. Statist. Assoc.* **77** 605–610. With comments by Joseph B. Kadane and D. V. Lindley and a rejoinder by the author. MR0675887

DE FINETTI, B. (1937). Foresight: Its logical laws, its subjective sources. In *Breakthroughs in Statistics* 134–174. Springer, New York.

DECK, K. M., PAYNE, M. and HOLMAN, M. J. (2013). First-order resonance overlap and the stability of close two-planet systems. *Astrophys. J.* **774** 129. https://doi.org/10.1088/0004-637X/774/2/129

DELISLE, J. B., HARA, N. and SÉGRANSAN, D. (2020a). Efficient modeling of correlated noise. I. Statistical significance of periodogram peaks. *A & A* **635** A83. https://doi.org/10.1051/0004-6361/201936905

DELISLE, J. B., HARA, N. and SÉGRANSAN, D. (2020b). Efficient modeling of correlated noise. II. A flexible noise model with fast and scalable methods. *A & A* **638** A95. https://doi.org/10.1051/0004-6361/201936906

DÍAZ, R. F., SÉGRANSAN, D., UDRY, S., LOVIS, C., PEPE, F., DUMUSQUE, X., MARMIER, M., ALONSO, R., BENZ, W. et al. (2016). The HARPS search for southern extra-solar planets. XXXVIII. Bayesian re-analysis of three systems. New super-earths, unconfirmed signals, and magnetic cycles. *A & A* **585**. https://doi.org/10.1051/0004-6361/201526729

DRAPER, D. (2006). Coherence and calibration: Comments on subjectivity and "objectivity" in Bayesian analysis (comment on articles by Berger and by Goldstein). *Bayesian Anal.* **1** 423–428. MR2221274 https://doi.org/10.1214/06-BA116B

EFRON, B., TIBSHIRANI, R., STOREY, J. D. and TUSHER, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96** 1151–1160. MR1946571 https://doi.org/10.1198/016214501753382129

ENGEL, P. (1998). Ramsey: Croyance, vérité et probabilité. In *Le Formalisme en Question* (*Saint-Malo*, 1994). *Probl. Controv.* 289–302. Vrin, Paris. MR1695987

FARIA, J. P., HAYWOOD, R. D., BREWER, B. J., FIGUEIRA, P., OSHAGH, M., SANTERNE, A. and SANTOS, N. C. (2016). Uncovering the planets and stellar activity of CoRoT-7 using only radial velocities. *A & A* **588** A31. https://doi.org/10.1051/0004-6361/201527899

FORD, E. B. and GREGORY, P. C. (2007). Bayesian model selection and extrasolar planet detection. In *Statistical Challenges in Modern Astronomy IV* (G. J. Babu and E. D. Feigelson, eds.). *Astronomical Society of the Pacific Conference Series* **371** 189. https://doi.org/10.48550/arXiv.astro-ph/0608328

FOREMAN-MACKEY, D., AGOL, E., AMBIKASARAN, S. and ANGUS, R. (2017). Fast and scalable Gaussian process modeling with applications to astronomical time series. *Astron. J.* **154** 220. https://doi.org/10.3847/1538-3881/aa9332

FOSTER, D. P. and VOHRA, R. V. (1998). Asymptotic calibration. *Biometrika* **85** 379–390. MR1649119 https://doi.org/10.1093/biomet/85.2.379

GASCHEAU, M. (1843). Examen d'une classe d'équations différentielles et applicationa un cas particulier du probleme des trois corps. *C. R. Séances Acad. Sci.* **16** 393.

GELMAN, A., MENG, X.-L. and STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica* **6** 733–807. With comments and a rejoinder by the authors. MR1422404

GERINGER-SAMETH, A., WALKER, M. G., KOUSHIAPPAS, S. M., KOPOSOV, S. E., BELOKUROV, V., TORREALBA, G. and EVANS, N. W. (2015). Indication of gamma-ray emission from the newly discovered Dwarf Galaxy Reticulum II. *Phys. Rev. Lett.* **115** 081101. https://doi.org/10.1103/PhysRevLett.115.081101

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

GREGORY, P. C. (2007). A Bayesian periodogram finds evidence for three planets in HD 11964. *Mon. Not. R. Astron. Soc.* **381** 1607–1616. https://doi.org/10.1111/j.1365-2966.2007.12361.x

GUINDANI, M., MÜLLER, P. and ZHANG, S. (2009). A Bayesian discovery procedure. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 905–925. MR2750250 https://doi.org/10.1111/j.1467-9868.2009.00714.x

HANDLEY, W. J., HOBSON, M. P. and LASENBY, A. N. (2015a). polychord: Nested sampling for cosmology. *Mon. Not. R. Astron. Soc.* **450** L61–L65. https://doi.org/10.1093/mnrasl/slv047

HANDLEY, W. J., HOBSON, M. P. and LASENBY, A. N. (2015b). POLYCHORD: Next-generation nested sampling. *Mon. Not. R. Astron. Soc.* **453** 4384–4398. https://doi.org/10.1093/mnras/stv1911

HARA, N. C., BOUÉ, G., LASKAR, J. and CORREIA, A. C. M. (2017). Radial velocity data analysis with compressed sensing techniques. *Mon. Not. R. Astron. Soc.* **464** 1220–1246. https://doi.org/10.1093/mnras/stw2261

HARA, N. C., BOUÉ, G., LASKAR, J., DELISLE, J. B. and UNGER, N. (2019). Bias and robustness of eccentricity estimates from radial velocity data. *Mon. Not. R. Astron. Soc.* **489** 738–762. https://doi.org/10.1093/mnras/stz1849

HARA, N. C., DE POYFERRÉ, T., DELISLE, J.-B. and HOFFMANN, M. (2024). Supplement to "A continuous multiple hypothesis testing framework for optimal exoplanet detection." https://doi.org/10.1214/23-AOAS1810SUPPA, https://doi.org/10.1214/23-AOAS1810SUPPB

HARA, N. C., DELISLE, J. and UNGER, N. (2022). Testing whether a signal is strictly periodic. Application to disentangling planets and stellar activity in radial velocities. *A & A* **658** A177. https://doi.org/10.1051/0004-6361/202141197

HARA, N. C. and FORD, E. B. (2023). Statistical methods for exoplanet detection with radial velocities. *Annu. Rev. Stat. Appl.* **10** 623–649. MR4567808 https://doi.org/10.1146/annurev-statistics-033021-012225

HARA, N. C., UNGER, N., DELISLE, J. and DÍAZ, R. F. (2022). Detecting exoplanets with the false inclusion probability. Comparison with other detection criteria in the context of radial velocities. *A & A* **663**. https://doi.org/10.1051/0004-6361/202140543

JAYNES, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge Univ. Press, Cambridge. Edited and with a foreword by G. Larry Bretthorst. MR1992316 https://doi.org/10.1017/CBO9780511790423

LELEU, A., ROBUTEL, P. and CORREIA, A. C. M. (2015). Detectability of quasi-circular co-orbital planets. Application to the radial velocity technique. *A & A* **581** A128. https://doi.org/10.1051/0004-6361/201526175

LOVIS, C., MAYOR, M., PEPE, F., ALIBERT, Y., BENZ, W., BOUCHY, F. and CORREIA, A. C. M. (2006). An extrasolar planetary system with three Neptune-mass planets. *Nature* **441** 305–309. https://doi.org/10.1038/nature04828

LOVIS, C., SÉGRANSAN, D., MAYOR, M., UDRY, S., BENZ, W., BERTAUX, J. L., BOUCHY, F., CORREIA, A. C. M., LASKAR, J. et al. (2011). The HARPS search for southern extra-solar planets. XXVIII. Up to seven planets orbiting HD 10180: Probing the architecture of low-mass planetary systems. *A & A* **528** A112. https://doi.org/10.1051/0004-6361/201015577

MÜLLER, P., PARMIGIANI, G. and RICE, K. (2007). FDR and Bayesian multiple comparisons rules. In *Bayesian Statistics* 8. *Oxford Sci. Publ.* 349–370. Oxford Univ. Press, Oxford. MR2433200

MÜLLER, P., PARMIGIANI, G., ROBERT, C. and ROUSSEAU, J. (2004). Optimal sample size for multiple testing: The case of gene expression microarrays. *J. Amer. Statist. Assoc.* **99** 990–1001. MR2109489 https://doi.org/10.1198/016214504000001646

NELSON, B. E., FORD, E. B., BUCHNER, J., CLOUTIER, R. and DÍAZ, R. F. (2020). Quantifying the Bayesian evidence for a planet in radial velocity data. *Astron. J.* **159** 73. https://doi.org/10.3847/1538-3881/ab5190

PERRYMAN, M. (2018). *The Exoplanet Handbook*.

REDNER, R. A. and WALKER, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* **26** 195–239. MR0738930 https://doi.org/10.1137/1026034

ROODAKI, A., BECT, J. and FLEURY, G. (2014). Relabeling and summarizing posterior distributions in signal decomposition problems when the number of components is unknown. *IEEE Trans. Signal Process.* **62** 4091–4104. MR3260412 https://doi.org/10.1109/TSP.2014.2333569

RUBIN, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.* **12** 1151–1172. MR0760681 https://doi.org/10.1214/aos/1176346785

SCOTT, J. G. and BERGER, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *J. Statist. Plann. Inference* **136** 2144–2162. MR2235051 https://doi.org/10.1016/j.jspi.2005.08.031

SKILLING, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Anal.* **1** 833–859. MR2282208 https://doi.org/10.1214/06-BA127

SONG, H., DIETHE, T. and KULL, M. (2019). Distribution calibration for regression. In *Proceedings of the* 36*th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.). *Proceedings of Machine Learning Research* **97** 5897–5906. PMLR.

STEPHENS, M. (2000). Dealing with label switching in mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 795–809. MR1796293 https://doi.org/10.1111/1467-9868.00265

STEPHENS, M. (2017). False discovery rates: A new deal. *Biostatistics* **18** 275–294. MR3824755 https://doi.org/10.1093/biostatistics/kxw041

STOREY, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Statist.* **31** 2013–2035. MR2036398 https://doi.org/10.1214/aos/1074290335

TITTERINGTON, D. M., SMITH, A. F. M. and MAKOV, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, Chichester. MR0838090

TOULIS, P. and BEAN, J. (2021). Randomization inference of periodicity in unequally spaced time series with application to exoplanet detection. arXiv e-prints. Available at arXiv:2105.14222.

TUOMI, M. and KOTIRANTA, S. (2009). Bayesian analysis of the radial velocities of HD 11506 reveals another planetary companion. *A & A* **496** L13–L16. https://doi.org/10.1051/0004-6361/200811531

UDRY, S., DUMUSQUE, X., LOVIS, C., SÉGRANSAN, D., DIAZ, R. F., BENZ, W., BOUCHY, F., COFFINET, A., LO CURTO, G. et al. (2019). The HARPS search for southern extra-solar planets. XLIV. Eight HARPS multi-planet systems hosting 20 super-Earth and Neptune-mass companions. *A & A* **622** A37. https://doi.org/10.1051/0004-6361/201731173

UNGER, N., SÉGRANSAN, D., QUELOZ, D., UDRY, S., LOVIS, C., MORDASINI, C., AHRER, E., BENZ, W., BOUCHY, F. et al. (2021). The HARPS search for southern extra-solar planets. XLVI. 12 super-earths around the solar type stars HD 39194, HD 93385, HD 96700, HD 154088, and HD 189567. *A & A* **654** A104. https://doi.org/10.1051/0004-6361/202141351

VEHTARI, A. and LAMPINEN, J. (2002). Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Comput.* **14** 2439–2468. https://doi.org/10.1162/08997660260293292

VON NEUMANN, J. and MORGENSTERN, O. (1947). *Theory of Games and Economic Behavior*. 2nd ed. Princeton Univ. Press, Princeton, NJ. MR0021298

VOVK, V. and SHAFER, G. (2005). Good randomized sequential probability forecasting is always possible. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 747–763. MR2210691 https://doi.org/10.1111/j.1467-9868.2005.00525.x

WRIGHT, J. T. and HOWARD, A. W. (2009). Efficient fitting of multiplanet Keplerian models to radial velocity and astrometry data. *Astrophys. J., Suppl. Ser.* **182** 205–215. https://doi.org/10.1088/0067-0049/182/1/205

ZECHMEISTER, M. and KÜRSTER, M. (2009). The generalised Lomb–Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. *A & A* **496** 577–584. https://doi.org/10.1051/0004-6361:200811296

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 770–793
https://doi.org/10.1214/23-AOAS1812

DENSITY-BASED MATCHING RULE: OPTIMALITY, ESTIMATION, AND APPLICATION IN FORENSIC PROBLEMS

BY HANA LEE[1,a], YUMOU QIU[2,d], ALICIA CARRIQUIRY[1,b] AND DANICA OMMEN[1,c]

[1]*Department of Statistics and Center for Statistics and Applications in Forensic Evidence, Iowa State University,* [a]*hnlee@iastate.edu,* [b]*alicia@iastate.edu,* [c]*dmommen@iastate.edu*

[2]*School of Mathematical Sciences and Center for Statistical Science, Peking University,* [d]*qiuyumou@math.pku.edu.cn*

We consider matching problems where the goal is to determine whether two observations randomly drawn from a population with multiple (sub)groups are from the same (sub)group. This is a key question in forensic science, where items with unidentified origins from suspects and crime scenes are compared to objects from a known set of sources to see if they originated from the same source. We derive the optimal matching rule under known density functions of data that minimizes the decision error probabilities. Empirically, the proposed matching rule is computed by plugging parametrically estimated density functions using training data into the formula of the optimal matching rule. The connections between the optimal matching rule and existing methods in forensic science are explained. In particular, we contrast the optimal matching rule to classification and also compare it to a score-based approach that relies on similarity features extracted from paired items. Numerical simulations are conducted to evaluate the proposed method and show that it outperforms the existing methods in terms of a higher ROC curve and higher power to identify matched pairs of items. We also demonstrate the utility of the proposed method by applying it to a real forensic data analysis of glass fragments.

REFERENCES

AITKEN, C. G. G. and LUCY, D. (2004). Evaluation of trace evidence in the form of multivariate data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **53** 109–122.

AITKEN, C. G. G., TARONI, F. and BOZZA, S. (2020). *Statistics and the Evaluation of Evidence for Forensic Scientists*. Wiley, Hoboken, NJ.

ALPAYDIN, E. (2020). *Introduction to Machine Learning*. MIT Press.

ANTHONIOZ, N. M. E. and CHAMPOD, C. (2014). Evidence evaluation in fingerprint comparison and automated fingerprint identification systems—modeling between finger variability. *Forensic Sci. Int.* **235** 86–101.

ATHEY, S., TIBSHIRANI, J. and WAGER, S. (2019). Generalized random forests. *Ann. Statist.* **47** 1148–1178.

AZZALINI, A. and CAPITANIO, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 579–602.

AZZALINI, A. and VALLE, A. D. (1996). The multivariate skew-normal distribution. *Biometrika* **83** 715–726.

BABIĆ, S., GELBGRAS, L., HALLIN, M. and LEY, C. (2021). Optimal tests for elliptical symmetry: Specified and unspecified location. *Bernoulli* **27** 2189–2216.

BERGER, J. O. (2013). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York.

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

BUTLER, J. M., IYER, H. K., PRESS, R. A., TAYLOR, M. K., VALLONE, P. M. and WILLIS, S. (2020). NIST scientific foundation reviews.

CARRIQUIRY, A., HOFMANN, H., TAI, X. H. and VANDERPLAS, S. (2019). Machine learning in forensic applications. *Significance* **16** 29–35.

COOK, R., EVETT, I. W., JACKSON, G., JONES, P. J. and LAMBERT, J. A. (1998). A hierarchy of propositions: Deciding which level to address in casework. *Sci. Justice* **38** 231–239.

EGLI, N. M., CHAMPOD, C. and MARGOT, P. (2007). Evidence evaluation in fingerprint comparison and automated fingerprint identification systems—modelling within finger variability. *Forensic Sci. Int.* **167** 189–195. https://doi.org/10.1016/j.forsciint.2006.06.054

Key words and phrases. Classification, forensic science, likelihood ratio, matching, optimal decision rule.

FAN, J. and FAN, Y. (2008). High dimensional classification using features annealed independence rules. *Ann. Statist.* **36** 2605.

FAN, J., FENG, Y., JIANG, J. and TONG, X. (2016). Feature augmentation via nonparametrics and selection (FANS) in high-dimensional classification. *J. Amer. Statist. Assoc.* **111** 275–287.

HARE, E., HOFMANN, H. and CARRIQUIRY, A. (2017). Automatic matching of bullet land impressions. *Ann. Appl. Stat.* **11** 2332–2356.

HEPLER, A. B., SAUNDERS, C. P., DAVIS, L. J. and BUSCAGLIA, J. (2012). Score-based likelihood ratios for handwriting evidence. *Forensic Sci. Int.* **219** 129–140. https://doi.org/10.1016/j.forsciint.2011.12.009

JOHNSON, M. Q. and OMMEN, D. M. (2022). Handwriting identification using random forests and score-based likelihood ratios. *Stat. Anal. Data Min. ASA Data Sci. J.* **15** 357–375.

JOHNSON, R. A. and WICHERN, D. W. (2002). *Applied Multivariate Statistical Analysis* 5. Prentice Hall, Upper Saddle River, NJ.

KIBRIA, B. G. and JOARDER, A. H. (2006). A short review of multivariate t-distribution. *J. Statist. Res.* **40** 59–72.

KORKMAZ, S., GÖKSÜLÜK, D. and ZARARSIZ, G. (2014). MVN: An R package for assessing multivariate normality. *R J.* **6**.

KWAN, Q. Y. (1977). *Inference of Identity of Source*. Univ. California, Berkeley, CA.

LEE, H., QIU, Y. CARRIQUIRY, A. and OMMEN, D. (2024). Supplement to "Density-based matching rule: Optimality, estimation, and application in forensic problems." https://doi.org/10.1214/23-AOAS1812SUPPA, https://doi.org/10.1214/23-AOAS1812SUPPB

LEY, C., BABIĆ, S. and PALANGETIĆ, M. (2021). Elliptical symmetry tests in R. *R J.* **13**.

LINDLEY, D. V. (1977). A problem in forensic science. *Biometrika* **64** 207–213.

MARDIA, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya*, *Ser. B* 115–128.

MCANDREW, W. P. and HOUCK, M. M. (2020). Interpol review of forensic science management literature 2016–2019. *Forensic Sci. Int. Synerg.* **2** 382–388. https://doi.org/10.1016/j.fsisyn.2020.01.007

MCLACHLAN, G. J., LEE, S. X. and RATHNAYAKE, S. I. (2019). Finite mixture models. *Annu. Rev. Stat. Appl.* **6** 355–378.

MORRISON, G. S. and ENZINGER, E. (2018). Score based procedures for the calculation of forensic likelihood ratios–scores should take account of both similarity and typicality. *Sci. Justice* **58** 47–58.

NEUMANN, C. and AUSDEMORE, M. (2020). Defence against the modern arts: The curse of statistics—part II: 'score-based likelihood ratios'. *Law Probab. Risk* **19** 21–42.

OMMEN, D. M. and SAUNDERS, C. P. (2021). A problem in forensic science highlighting the differences between the Bayes factor and likelihood ratio. *Statist. Sci.* **36** 344–359.

PARK, S. and CARRIQUIRY, A. (2019). Learning algorithms to evaluate forensic glass evidence. *Ann. Appl. Stat.* **13** 1068–1102.

PARK, S., CARRIQUIRY, A., HORKLEY, L. K. and PEATE, D. W. (2020). A database of elemental compositions of architectural float glass samples measured by LA-ICP-MS. *Data Brief* **30** 105449. https://doi.org/10.1016/j.dib.2020.105449

REINDERS, S., GUAN, Y., OMMEN, D. and NEWMAN, J. (2022). Source-anchored, trace-anchored, and general match score-based likelihood ratios for camera device identification. *J. Forensic Sci.* **67** 975–988. https://doi.org/10.1111/1556-4029.14991

TREJOS, T., KOCH, S. and MEHLTRETTER, A. (2020). Scientific foundations and current state of trace evidence—a review. *Forensic Chem.* **18** 100223.

VENERI, F. and OMMEN, D. (2022). Ensemble of SLR systems for forensic evidence. Poster presented at the 74th Annual Scientific Conference of the American Academy of Forensic Sciences (AAFS), Seattle, Washington.

WAGER, S. and ATHEY, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* **113** 1228–1242.

WAHBA, G. (2002). Soft and hard classification by reproducing kernel Hilbert space methods. *Proc. Natl. Acad. Sci. USA* **99** 16524–16530.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 794–818
https://doi.org/10.1214/23-AOAS1813

LAND-USE FILTERING FOR NONSTATIONARY SPATIAL PREDICTION OF COLLECTIVE EFFICACY IN AN URBAN ENVIRONMENT

BY J. BRANDON CARTER[1,a], CHRISTOPHER R. BROWNING[2,c], BETHANY BOETTNER[3,e], NICOLO PINCHAK[2,d] AND CATHERINE A. CALDER[1,b]

[1]*Department of Statistics and Data Sciences, University of Texas at Austin,* [a]*carterjb@utexas.edu,* [b]*calder@austin.utexas.edu*
[2]*Department of Sociology, The Ohio State University,* [c]*browning.90@osu.edu,* [d]*pinchak.5@osu.edu*
[3]*Population Research Institute, Ohio State University,* [e]*boettner.6@osu.edu*

Collective efficacy—the capacity of communities to exert social control toward the realization of their shared goals—is a foundational concept in the urban sociology and neighborhood effects literature. Traditionally, empirical studies of collective efficacy use large sample surveys to estimate collective efficacy of different neighborhoods within an urban setting. Such studies have demonstrated an association between collective efficacy and local variation in community violence, educational achievement, and health. Unlike traditional collective efficacy measurement strategies, the Adolescent Health and Development in Context (AHDC) Study implemented a new approach, obtaining spatially-referenced, place-based ratings of collective efficacy from a representative sample of individuals residing in Columbus, OH. In this paper we introduce a novel nonstationary spatial model for interpolation of the AHDC collective efficacy ratings across the study area, which leverages administrative data on land use. Our constructive model specification strategy involves dimension expansion of a latent spatial process and the use of a filter defined by the land-use partition of the study region to connect the latent multivariate spatial process to the observed ordinal ratings of collective efficacy. Careful consideration is given to the issues of parameter identifiability, computational efficiency of an MCMC algorithm for model fitting, and fine-scale spatial prediction of collective efficacy.

REFERENCES

ALBERT, J. and CHIB, S. (1997). Bayesian methods for cumulative, sequential and two-step ordinal data regression models.

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

APANASOVICH, T. V., GENTON, M. G. and SUN, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *J. Amer. Statist. Assoc.* **107** 180–193. MR2949350 https://doi.org/10.1080/01621459.2011.643197

FRANKLIN COUNTY AUDITOR (2021). Franklin county GIS parcel shape files. 2014 Data retrieved from https://franklincountyauditor.com/ftp.

BANDURA, A. (1982). Self-efficacy mechanism in human agency. *Amer. Psychol.* **37** 122–147.

BANDURA, A. (1986). *Social Foundations of Thought and Action*: *A Social Cognitive Theory*. *Prentice-Hall Series in Social Learning Theory*. Prentice-Hall, Englewood Cliffs, NJ.

BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. MR3362184

BERRETT, C. and CALDER, C. A. (2012). Data augmentation strategies for the Bayesian spatial probit regression model. *Comput. Statist. Data Anal.* **56** 478–490. MR2853748 https://doi.org/10.1016/j.csda.2011.08.020

BERRETT, C. and CALDER, C. A. (2016). Bayesian spatial binary classification. *Spat. Stat.* **16** 72–102. MR3493089 https://doi.org/10.1016/j.spasta.2016.01.004

Key words and phrases. Bayesian statistics, data augmentation, dimension expansion, nonstationarity, sociology, spatial statistics.

BESANÇON, M., PAPAMARKOU, T., ANTHOFF, D., ARSLAN, A., BYRNE, S., LIN, D. and PEARSON, J. (2021). Distributions.jl: Definition and modeling of probability distributions in the JuliaStats ecosystem. *J. Stat. Softw.* **98** 1–30. https://doi.org/10.18637/jss.v098.i16

BEZANSON, J., EDELMAN, A., KARPINSKI, S. and SHAH, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Rev.* **59** 65–98. MR3605826 https://doi.org/10.1137/141000671

BORNN, L., SHADDICK, G. and ZIDEK, J. V. (2012). Modeling nonstationary processes through dimension expansion. *J. Amer. Statist. Assoc.* **107** 281–289. MR2949359 https://doi.org/10.1080/01621459.2011.646919

BROWNING, C. R. (2002). The span of collective efficacy: Extending social disorganization theory to partner violence. *J. Marriage Fam.* **64** 833–850.

CALDER, C. A. (2008). A dynamic process convolution approach to modeling ambient particulate matter concentrations. *Environmetrics* **19** 39–48. MR2416543 https://doi.org/10.1002/env.852

CARTER, J. B., BROWNING, C. R., BOETTNER, B., PINCHAK, N. and CALDER, C. A. (2024). Supplement to "Land-use filtering for nonstationary spatial prediction of collective efficacy in an urban environment." https://doi.org/10.1214/23-AOAS1813SUPP

COHEN, D. A., FINCH, B. K., BOWER, A. and SASTRY, N. (2006). Collective efficacy and obesity: The potential influence of social factors on health. *Soc. Sci. Med.* **62** 769–778.

COLEMAN, J. S. (1988). Social capital in the creation of human capital. *Am. J. Sociol.* **94** S95–S120.

CORCORAN, J., ZAHNOW, R., WICKES, R. and HIPP, J. (2018). Neighbourhood land use features, collective efficacy and local civic actions. *Urban Stud.* **55** 2372–2390.

COWLES, M. K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Stat. Comput.* **6** 101–111.

DE OLIVEIRA, V. (1997). *Prediction in Some Classes of Non-Gaussian Random Fields*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—University of Maryland, College Park. MR2696531

DE OLIVEIRA, V. (2000). Bayesian prediction of clipped Gaussian random fields. *Comput. Statist. Data Anal.* **34** 299–314.

FRIGO, M. and JOHNSON, S. G. (2005). The design and implementation of FFTW3. *Proc. IEEE* **93** 216–231. Special issue on "Program Generation, Optimization, and Platform Adaptation". https://doi.org/10.1109/JPROC.2004.840301

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

GNEITING, T., KLEIBER, W. and SCHLATHER, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc.* **105** 1167–1177. MR2752612 https://doi.org/10.1198/jasa.2010.tm09420

GOULARD, M. and VOLTZ, M. (1992). Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. *Math. Geol.* **24** 269–286.

GRZEBYK, M. and WACKERNAGEL, H. (1994). Multivariate analysis and spatial/temporal scales: Real and complex models. In *Proceedings of the XVIIth International Biometrics Conference* **1** 19–33. Citeseer.

GUSTAFSON, P. (2015). *Bayesian Inference for Partially Identified Models*: *Exploring the Limits of Limited Data*. *Monographs on Statistics and Applied Probability* **141**. CRC Press, Boca Raton, FL. MR3642458

HIGGS, M. D. and HOETING, J. A. (2010). A clipped latent variable model for spatially correlated ordered categorical data. *Comput. Statist. Data Anal.* **54** 1999–2011. MR2640303 https://doi.org/10.1016/j.csda.2010.02.024

HIPP, J. R. (2016). Collective efficacy: How is it conceptualized, how is it measured, and does it really matter for understanding perceived neighborhood crime and disorder? *J. Crim. Justice* **46** 32–44.

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

JACOBS, J. (1961). *The Death and Life of Great American Cities*. Random House, New York.

LI, F., DING, P. and MEALLI, F. (2023). Bayesian causal inference: A critical review. *Philos. Trans. R. Soc. Lond. A* **381** Paper No. 20220153, 24 pp. MR4590497 https://doi.org/10.1098/rsta.2022.0153

LIN, D., WHITE, J. M., BYRNE, S., BATES, D., NOACK, A., PEARSON, J., ARSLAN, A., SQUIRE, K., ANTHOFF, D. et al. (2019). JuliaStats/Distributions.jl: A Julia package for probability distributions and associated functions. https://doi.org/10.5281/zenodo.2647458

LINERO, A. R., BRADLEY, J. R. and DESAI, A. (2018). Multi-rubric models for ordinal spatial data with application to online ratings data. *Ann. Appl. Stat.* **12** 2054–2074. MR3875692 https://doi.org/10.1214/18-AOAS1143

MATSUEDA, R. L. and DRAKULICH, K. M. (2016). Measuring collective efficacy: A multilevel measurement model for nested data. *Sociol. Methods Res.* **45** 191–230. MR3484271 https://doi.org/10.1177/0049124115578030

MOGENSEN, P. K. and RISETH, A. N. (2018). Optim: A mathematical optimization package for Julia. *J. Open Sour. Softw.* **3** 615. https://doi.org/10.21105/joss.00615

MOLNAR, B. E., GOERGE, R. M., GILSANZ, P., HILL, A., SUBRAMANIAN, S. V., HOLTON, J. K., DUNCAN, D. T., BEATRIZ, E. D. and BEARDSLEE, W. R. (2016). Neighborhood-level social processes and substantiated cases of child maltreatment. *Child Abuse Neglect* **51** 41–53. https://doi.org/10.1016/j.chiabu.2015.11.007

NETO, J. H. V., SCHMIDT, A. M. and GUTTORP, P. (2014). Accounting for spatially varying directional effects in spatial covariance structures. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **63** 103–122. MR3148271 https://doi.org/10.1111/rssc.12027

PEBESMA, E. (2018). Simple features for R: Standardized support for spatial vector data. *R J.* **10** 439–446. https://doi.org/10.32614/RJ-2018-009

R CORE TEAM (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

RACKAUCKAS, C. and NIE, Q. (2017). DifferentialEquations.jl—A performant and feature-rich ecosystem for solving differential equations in Julia. *J. Open Res. Softw.* **5**. Exported from https://app.dimensions.ai on 2019/05/05. https://doi.org/10.5334/jors.151

RAUDENBUSH, S. W. and SAMPSON, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociol. Method.* **29** 1–41.

REICH, B. J., EIDSVIK, J., GUINDANI, M., NAIL, A. J. and SCHMIDT, A. M. (2011). A class of covariate-dependent spatiotemporal covariance functions. *Ann. Appl. Stat.* **5** 2265–2687.

RISSER, M. D. and CALDER, C. A. (2015). Regression-based covariance functions for nonstationary spatial modeling. *Environmetrics* **26** 284–297. MR3340964 https://doi.org/10.1002/env.2336

RISSER, M. D., CALDER, C. A., BERROCAL, V. J. and BERRETT, C. (2019). Nonstationary spatial prediction of soil organic carbon: Implications for stock assessment decision making. *Ann. Appl. Stat.* **13** 165–188. MR3937425 https://doi.org/10.1214/18-AOAS1204

SAMPSON, R. J., MORENOFF, J. D. and EARLS, F. (1999). Beyond social capital: Spatial dynamics of collective efficacy for children. *Am. Sociol. Rev.* **64** 633–660.

SAMPSON, R. J., RAUDENBUSH, S. W. and EARLS, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* **277** 918–924. https://doi.org/10.1126/science.277.5328.918

SCHLIEP, E. M. and HOETING, J. A. (2015). Data augmentation and parameter expansion for independent or spatially correlated ordinal data. *Comput. Statist. Data Anal.* **90** 1–14. MR3354825 https://doi.org/10.1016/j.csda.2015.03.020

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

SCHMIDT, A. M., GUTTORP, P. and O'HAGAN, A. (2011). Considering covariates in the covariance structure of spatial processes. *Environmetrics* **22** 487–500. MR2843403 https://doi.org/10.1002/env.1101

SHAW, C. R. and MCKAY, H. D. (1942). *Juvenile Delinquency and Urban Areas*. Univ. Chicago Press, Chicago, IL.

WACKERNAGEL, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*, 3rd ed. Springer, Berlin;.

WATANABE, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11** 3571–3594. MR2756194

WEISBURD, D., ECK, J. E., BRAGA, A. A., TELEP, C. W., CAVE, B., BOWERS, K., BRUINSMA, G., GILL, C., GROFF, E. R. et al. (2016). *Place Matters: Criminology for the Twenty-First Century*. Cambridge Univ. Press, Cambridge. https://doi.org/10.1017/CBO9781139342087

WICKES, R., ZAHNOW, R., CORCORAN, J. and HIPP, J. R. (2019). Neighbourhood social conduits and resident social cohesion. *Urban Stud.* **56** 226–248.

WICKHAM, H., AVERICK, M., BRYAN, J., CHANG, W., MCGOWAN, L. D., FRANÇOIS, R., GROLEMUND, G., HAYES, A., HENRY, L. et al. (2019). Welcome to the tidyverse. *J. Open Sour. Softw.* **4** 1686. https://doi.org/10.21105/joss.01686

YAN, J., COWLES, M. K., WANG, S. and ARMSTRONG, M. P. (2007). Parallelizing MCMC for Bayesian spatiotemporal geostatistical models. *Stat. Comput.* **17** 323–335. MR2409797 https://doi.org/10.1007/s11222-007-9022-2

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 819–840
https://doi.org/10.1214/23-AOAS1814

TESTING FOR THE CAUSAL MEDIATION EFFECTS OF MULTIPLE MEDIATORS USING THE KERNEL MACHINE DIFFERENCE METHOD IN GENOME-WIDE EPIGENETIC STUDIES

BY JINCHENG SHEN[1,a], JOEL SCHWARTZ[2,b], ANDREA A. BACCARELLI[3,c] AND XIHONG LIN[4,d]

[1] *Department of Population Health Sciences, University of Utah School of Medicine,* [a]*jincheng.shen@hsc.utah.edu*
[2] *Departments of Environmental Health, Harvard T.H. Chan School of Public Health,* [b]*jschwrtz@hsph.harvard.edu*
[3] *Departments of Environmental Health Sciences and Epidemiology, Columbia University, Mailman School of Public Health,* [c]*ab4303@cumc.columbia.edu*
[4] *Department of Biostatistics, Harvard T. H. Chan School of Public Health,* [d]*xlin@hsph.harvard.edu*

The rapid growth of high-throughput genomic and epigenomic data enables exploration of biological mechanisms underlying diseases causing processes beyond traditional association studies. Using the causal mediation analysis framework, we develop the kernel machine difference (KMD) method, which provides a testing procedure for detecting the mediation effects of a set of mediators, for example, the DNA methylation probes within a region or a gene. Our method extends the difference method in single mediator analysis to jointly model the mediatory role of the methylation of multiple neighboring probes, as they often work together in a collaborative fashion. Kernel machine regression is employed to accommodate flexible parametric and nonparametric effects of multiple mediators on the outcome and to allow for robust testing for the joint natural indirect effect (NIE). The proposed testing procedure does not require explicit modeling of the dependence of multiple mediators on exposure and confounders and the correlation among multiple mediators. It hence provides a robust and computationally efficient tool, especially for genomic regions with moderate to high-dimensional probes. We evaluate the performance of the proposed test using extensive simulations and demonstrate its gain in robustness and power when the effects are nonlinear. We apply the proposed test to the analysis of the epigenome-wide Normative Aging Study (NAS) to investigate the mediatory role of DNA methylation in the causal pathway between smoking behavior and lung function.

REFERENCES

AVIN, C., SHPITSER, I. and PEARL, J. (2005). Identifiability of path-specific effects. In *Proceedings of the* 19*th International Joint Conference on Artificial Intelligence* 357–363. Morgan Kaufmann, San Mateo.

BARON, R. M. and KENNY, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* **51** 1173–1182. https://doi.org/10.1037//0022-3514.51.6.1173

BELL, B., ROSE, C. L. and DAMON, A. (1972). The normative aging study: An interdisciplinary and longitudinal study of health and aging. *Int. J. Aging Hum. Dev.* **3** 5–17.

BIBIKOVA, M., BARNES, B., TSAN, C., HO, V., KLOTZLE, B., LE, J. M., DELANO, D., ZHANG, L., SCHROTH, G. P. et al. (2011). High density DNA methylation array with single CpG site resolution. *Genomics* **98** 288–295.

BOCK, C. (2012). Analysing and interpreting DNA methylation data. *Nat. Rev. Genet.* **13** 705–719. https://doi.org/10.1038/nrg3273

CLOGG, C. C., PETKOVA, E. and CHENG, T. (1995). Reply to Allison: More on comparing regression coefficients. *Amer. J. Sociol.* **100** 1305–1312.

COHEN, J. and COHEN, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Erlbaum, Hillsdale, NJ.

Key words and phrases. DNA methylation, difference method, hypothesis testing, kernel machine regression, mediation effects, multiple mediators, semiparametric regression.

DANIEL, R. M., DE STAVOLA, B. L., COUSENS, S. N. and VANSTEELANDT, S. (2015). Causal mediation analysis with multiple mediators. *Biometrics* **71** 1–14. MR3335344 https://doi.org/10.1111/biom.12248

GAO, X., JIA, M., ZHANG, Y., BREITLING, L. P. and BRENNER, H. (2015). DNA methylation changes of whole blood cells in response to active smoking exposure in adults: A systematic review of DNA methylation studies. *Clin. Epigenet.* **7** 113.

HAINMUELLER, J. and HAZLETT, C. (2013). Kernel regularized least squares: Reducing misspecification bias with a flexible and interpretable machine learning approach. *Polit. Anal.* mpt019.

HOUSEMAN, E. A., ACCOMANDO, W. P., KOESTLER, D. C., CHRISTENSEN, B. C., MARSIT, C. J., NELSON, H. H., WIENCKE, J. K. and KELSEY, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform.* **13** 1.

HUANG, Y.-T. and PAN, W.-C. (2016). Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics* **72** 402–413. MR3515767 https://doi.org/10.1111/biom.12421

IMAI, K., KEELE, L. and TINGLEY, D. (2010). A general approach to causal mediation analysis. *Psychol. Methods* **15** 309–334. https://doi.org/10.1037/a0020761

JIANG, Z. and VANDERWEELE, T. J. (2015). When is the difference method conservative for assessing mediation? *Amer. J. Epidemiol.* **182** 105–108. https://doi.org/10.1093/aje/kwv059

JOHNSON, W. E., LI, C. and RABINOVIC, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8** 118–127. https://doi.org/10.1093/biostatistics/kxj037

KIMELDORF, G. S. and WAHBA, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Stat.* **41** 495–502. MR0254999 https://doi.org/10.1214/aoms/1177697089

LEE, K. W. K. and PAUSOVA, Z. (2013). Cigarette smoking and DNA methylation. *Front. Genet.* **4** 132. https://doi.org/10.3389/fgene.2013.00132

LEPEULE, J., BACCARELLI, A., TARANTINI, L., MOTTA, V., CANTONE, L., LITONJUA, A. A., SPARROW, D., VOKONAS, P. S. and SCHWARTZ, J. (2012). Gene promoter methylation is associated with lung function in the elderly: The normative aging study. *Epigenetics* **7** 261–269.

LIN, D. Y., FLEMING, T. R. and DE GRUTTOLA, V. D. (1997). Estimating the proportion of treatment effect explained by a surrogate marker. *Stat. Med.* **16** 1515–1527. https://doi.org/10.1002/(sici)1097-0258(19970715)16:13<1515::aid-sim572>3.0.co;2-1

LIU, D., LIN, X. and GHOSH, D. (2007). Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models. *Biometrics* **63** 1079–1088, 1311. MR2414585 https://doi.org/10.1111/j.1541-0420.2007.00799.x

LOK, J. J. and BOSCH, R. J. (2021). Causal organic indirect and direct effects: Closer to the original approach to mediation analysis, with a product method for binary mediators. *Epidemiology* **32** 412–420.

MACKINNON, D. P. (2008). *Introduction to Statistical Mediation Analysis*. Routledge, London.

MACKINNON, D. P., LOCKWOOD, C. M., HOFFMAN, J. M., WEST, S. G. and SHEETS, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychol. Methods* **7** 83–104. https://doi.org/10.1037/1082-989x.7.1.83

MACKINNON, D. P., LOCKWOOD, C. M. and WILLIAMS, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivar. Behav. Res.* **39** 99–128.

MCGUIGAN, K. and LANGHOLTZ, B. (1988). A note on testing mediation paths using ordinary least-squares regression. Unpublished note.

OPITZ, C. A., LITZENBURGER, U. M., SAHM, F., OTT, M., TRITSCHLER, I., TRUMP, S., SCHUMACHER, T., JESTAEDT, L., SCHRENK, D. et al. (2011). An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor. *Nature* **478** 197–203.

PEARL, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* 411–420. Morgan Kaufmann, San Mateo.

QIU, W., BACCARELLI, A., CAREY, V. J., BOUTAOUI, N., BACHERMAN, H., KLANDERMAN, B., RENNARD, S., AGUSTI, A., ANDERSON, W. et al. (2012). Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. *Amer. J. Respir. Crit. Care Med.* **185** 373–381.

ROBINS, J. M. and GREENLAND, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3** 143–155. https://doi.org/10.1097/00001648-199203000-00013

ROBINS, J. M. and RICHARDSON, T. S. (2010). Alternative graphical causal models and the identification of direct effects. In *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures* 103–158.

SCHÖLKOPF, B. and SMOLA, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization. Optimization, and Beyond*. MIT Press.

SHEN, J., SCHWARTZ, J., BACCARELLI, A. A. and LIN, X. (2024). Supplement to “Testing for the causal mediation effects of multiple mediators using the kernel machine difference method in genome-wide epigenetic studies.” https://doi.org/10.1214/23-AOAS1814SUPP

SOBEL, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociol. Method.* **13** 290–312. MR2633309

SOBEL, M. E. (1987). Direct and indirect effects in linear structural equation models. *Sociol. Methods Res.* **16** 155–176.

STEEN, J. LOEYS, T., MOERKERKE, B. and VANSTEELANDT, S. (2017). Flexible mediation analysis with multiple mediators. *Amer. J. Epidemiol.* **186** 184–193. MR3888006

STENSRUD, M. J., HERNÁN, M. A., TCHETGEN TCHETGEN, E. J., ROBINS, J. M., DIDELEZ, V. and YOUNG, J. G. (2021). A generalized theory of separable effects in competing event settings. *Lifetime Data Anal.* **27** 588–631. MR4330432 https://doi.org/10.1007/s10985-021-09530-8

TAGURI, M., FEATHERSTONE, J. and CHENG, J. (2018). Causal mediation analysis with multiple causally non-ordered mediators. *Stat. Methods Med. Res.* **27** 3–19. MR3745651 https://doi.org/10.1177/0962280215615899

TCHETGEN, E. J. T. and SHPITSER, I. (2012). Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness and sensitivity analysis. *Ann. Statist.* **40** 1816–1845. MR3015045 https://doi.org/10.1214/12-AOS990

TRICHE, T. JR (2012). IlluminaHumanMethylation450k. db: Illumina Human Methylation 450k annotation data. R package version 1.

VANDERWEELE, T. J. and VANSTEELANDT, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Stat. Interface* **2** 457–468. MR2576399 https://doi.org/10.4310/SII.2009.v2.n4.a7

VANDERWEELE, T. J. and VANSTEELANDT, S. (2014). Mediation analysis with multiple mediators. *Epidemiol. Methods* **2** 95–115. https://doi.org/10.1515/em-2012-0010

VANSTEELANDT, S., BEKAERT, M. and LANGE, T. (2012). Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiol. Methods* **1** 131–158.

VAPNIK, V. N. (1998). *Statistical Learning Theory. Adaptive and Learning Systems for Signal Processing, Communications, and Control*. Wiley, New York. A Wiley-Interscience Publication. MR1641250

WANG, T., CHEN, M. and ZHAO, H. (2016). Estimating DNA methylation levels by joint modeling of multiple methylation profiles from microarray data. *Biometrics* **72** 354–363. MR3515762 https://doi.org/10.1111/biom.12422

ZHANG, H., ZHENG, Y., ZHANG, Z., GAO, T., JOYCE, B., YOON, G., ZHANG, W. and SCHWARTZ, J. (2016). Estimating and testing high-dimensional mediation effects in epigenetic studies. *Bioinformatics* **32** 3150–3154.

ZHENG, W. and VAN DER LAAN, M. J. (2012). Targeted maximum likelihood estimation of natural direct effects. *Int. J. Biostat.* **8** Art. 3, 42. MR2923277 https://doi.org/10.2202/1557-4679.1361

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 841–857
https://doi.org/10.1214/23-AOAS1815

WEAK SIGNAL INCLUSION UNDER DEPENDENCE AND APPLICATIONS IN GENOME-WIDE ASSOCIATION STUDY

BY X. JESSIE JENG[1,a], YIFEI HU[1,b], QUAN SUN[2,c] AND YUN LI[3,d]

[1]*Department of Statistics, North Carolina State University,* [a]*xjjeng@ncsu.edu,* [b]*hyifei@ncsu.edu*
[2]*Department of Biostatistics, University of North Carolina at Chapel Hill,* [c]*quansun@live.unc.edu*
[3]*Department of Biostatistics, Genetics, University of North Carolina at Chapel Hill,* [d]*yunli@med.unc.edu*

In this study we present a data-driven method called false negative control (FNC) screening to address the challenge of detecting weak signals in underpowered genome-wide association studies (GWASs), where true signals are often obscured by a large amount of noise. Our approach focuses on controlling false negatives and efficiently regulates the proportion of false negatives at a user-specified level in realistic settings with arbitrary covariance dependence between variables. We calibrate overall dependence using a parameter that aligns with the existing phase diagram in high-dimensional sparse inference, allowing us to asymptotically explicate the joint effect of covariance dependence, signal sparsity, and signal intensity on the proposed method. Our new phase diagram shows that FNC screening can efficiently select a set of candidate variables to retain a high proportion of signals, even when the signals are not individually separable from noise. We compare the performance of FNC screening to several existing methods in simulation studies, and the proposed method outperforms the others in adapting to a user-specified false negative control level. Moreover, we apply FNC screening to 145 GWAS datasets, obtained from the UK Biobank, and demonstrate a substantial increase in power to retain true signals for downstream analyses.

REFERENCES

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556.

ARIAS-CASTRO, E. and CHEN, S. (2017). Distribution-free multiple testing. *Electron. J. Stat.* **11** 1983–2001.

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.

CAI, T., JIN, J. and LOW, M. (2007). Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.* **35** 2421–2449.

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.

CAI, T. T. and JIN, J. (2010). Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. *Ann. Statist.* 100–145.

CAI, T. T. and SUN, W. (2017a). Large-scale global and simultaneous inference: Estimation and testing in very high dimensions. *Ann. Rev. Econ.* **9** 411–439.

CAI, T. T. and SUN, W. (2017b). Optimal screening and discovery of sparse signals with applications to multistage high-throughput studies. *J. R. Stat. Soc. Ser. B.* **79** 197–223.

CHEN, S. X., LI, J. and ZHONG, P.-S. (2019). Two-sample and ANOVA tests for high dimensional means. *Ann. Statist.* **47** 1443–1474.

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

DONOHO, D. and JIN, J. (2015). Special invited paper: Higher criticism for large-scale inference, especially for rare and weak effects. *Statist. Sci.* 1–25.

FAN, J., HAN, X. and GU, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *J. Amer. Statist. Assoc.* **107** 1019–1035.

Key words and phrases. Arbitrary covariance dependence, false negative control, underpowered GWAS, high-dimensional data, user-adaptive method.

Fowlkes, E. B. and Mallows, C. L. (1983). A method for comparing two hierarchical clusterings. *J. Amer. Statist. Assoc.* **78** 553–569.

Gao, Z. and Stoev, S. (2020). Fundamental limits of exact support recovery in high dimensions. *Bernoulli* **26** 2605–2638.

Gao, Z. and Stoev, S. (2021). *Concentration of Maxima and Fundamental Limits in High-Dimensional Testing and Inference*. Springer, Berlin.

Genovese, C. and Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 499–517.

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

Halkidi, M., Batistakis, Y. and Vazirgiannis, M. (2001). On clustering validation techniques. *J. Intell. Inf. Syst.* **17** 107–145.

Hu, Y., Stilp, A. M., McHugh, C. P., Rao, S., Jain, D., Zheng, X., Lane, J., Méric de Bellefon, S., Raffield, L. M. et al. (2021). Whole-genome sequencing association analysis of quantitative red blood cell phenotypes: The NHLBI TOPMed program. *Am. J. Hum. Genet.* **108** 874–893.

Huang, L., Rosen, J. D., Sun, Q., Chen, J., Wheeler, M. M., Zhou, Y., Min, Y.-I., Kooperberg, C., Conomos, M. P. et al. (2022). TOP-LD: A tool to explore linkage disequilibrium with TOPMed whole-genome sequence data. *Am. J. Hum. Genet.* **109** 1175–1181.

Ingster, Y. I. (1994). Minimax detection of a signal in? p metrics. *J. Math. Sci.* **68** 503–515.

Jeng, X. J. (2023). Estimating the proportion of signal variables under arbitrary covariance dependence. *Electron. J. Stat.* **17** 950–979.

Jeng, X. J. and Chen, X. (2019). Variable selection via adaptive false negative control in linear regression. *Electron. J. Stat.* **13** 5306–5333.

Jeng, X. J., Daye, Z. J., Lu, W. and Tzeng, J.-Y. (2016). Rare variants association analysis in large-scale sequencing studies at the single locus level. *PLoS Comput. Biol.* **12** e1004993. https://doi.org/10.1371/journal.pcbi.1004993

Jeng, X. J., Hu, Y., Sun, Q. and Li, Y. (2024). Supplement to "Weak signal inclusion under dependence and applications in genome-wide association study." https://doi.org/10.1214/23-AOAS1815SUPP

Jeng, X. J., Zhang, T. and Tzeng, J.-Y. (2019). Efficient signal inclusion with genomic applications. *J. Amer. Statist. Assoc.* **114** 1787–1799.

Ji, P. and Jin, J. (2012). UPS delivers optimal phase diagram in high-dimensional variable selection. *Ann. Statist.* **40** 73–103.

Ji, P. and Zhao, Z. (2014). Rate optimal multiple testing procedure in high-dimensional regression. ArXiv preprint. Available at arXiv:1404.2961.

Jin, J., Ke, Z. T. and Wang, W. (2017). Phase transitions for high dimensional clustering and related problems. *Ann. Statist.* **45** 2151–2189.

Liu, W., Sun, Q., Huang, L., Bhattacharya, A., Wang, G. W., Tan, X., Kuban, K. C. K., Joseph, R. M., O'Shea, T. M. et al. (2022). Innovative computational approaches shed light on genetic mechanisms underlying cognitive impairment among children born extremely preterm. *J. Neurodev. Disord.* **14** 16.

Mahajan, A., Spracklen, C. N., Zhang, W., Ng, M. C. Y., Petty, L. E., Kitajima, H., Yu, G. Z., Rüeger, S., Speidel, L. et al. (2022). Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat. Genet.* **54** 560–572.

Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J. A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M. et al. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53** 1097–1103.

Meinshausen, N. and Rice, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.* **34** 373–393.

Mikhaylova, A. V., McHugh, C. P., Polfus, L. M., Raffield, L. M., Boorgula, M. P., Blackwell, T. W., Brody, J. A., Broome, J., Chami, N. et al. (2021). Whole-genome sequencing in diverse subjects identifies genetic correlates of leukocyte traits: The NHLBI TOPMed program. *Am. J. Hum. Genet.* **108** 1836–1851.

Sarkar, S. K. (2006). False discovery and false nondiscovery rates in single-step multiple testing procedures. *Ann. Statist.* **34** 394–415.

Sun, Q., Graff, M., Rowland, B., Wen, J., Huang, L., Miller-Fleming, T. W., Haessler, J., Preuss, M. H., Chai, J.-F. et al. (2022). Analyses of biomarker traits in diverse UK biobank participants identify associations missed by European-centric analysis strategies. *J. Hum. Genet.* **67** 87–93.

Vuckovic, D., Bao, E. L., Akbari, P., Lareau, C. A., Mousas, A., Jiang, T., Chen, M.-H., Raffield, L. M., Tardaguila, M. et al. (2020). The polygenic and monogenic basis of blood traits and diseases. *Cell* **182** 1214–1231.e11.

WESTFALL, P. H. and YOUNG, S. S. (1993). *Resampling-Based Multiple Testing*: *Examples and Methods for P-Value Adjustment* **279**. Wiley, New York.

ZHAO, B., LI, T., SMITH, S. M., XIONG, D., WANG, X., YANG, Y., LUO, T., ZHU, Z., SHAN, Y. et al. (2022). Common variants contribute to intrinsic human brain functional networks. *Nat*. *Genet*. **54** 508–517.

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 858–881
https://doi.org/10.1214/23-AOAS1817

A SIMPLE AND FLEXIBLE TEST OF SAMPLE EXCHANGEABILITY WITH APPLICATIONS TO STATISTICAL GENOMICS

BY ALAN J. AW[1,a], JEFFREY P. SPENCE[2,b] AND YUN S. SONG[3,c]

[1]*Department of Statistics, University of California, Berkeley,* [a]*alanaw1@berkeley.edu*
[2]*Department of Genetics, School of Medicine, Stanford University,* [b]*jspence@stanford.edu*
[3]*Department of Statistics and Computer Science Division, University of California, Berkeley,* [c]*yss@berkeley.edu*

In scientific studies involving analyses of multivariate data, basic but important questions often arise for the researcher: Is the sample exchangeable, meaning that the joint distribution of the sample is invariant to the ordering of the units? Are the features independent of one another, or perhaps the features can be grouped so that the groups are mutually independent? In statistical genomics these considerations are fundamental to downstream tasks such as demographic inference and the construction of polygenic risk scores. We propose a nonparametric approach, which we call the V test, to address these two questions, namely, a test of sample exchangeability given dependency structure of features and a test of feature independence given sample exchangeability. Our test is conceptually simple, yet fast and flexible. It controls the Type I error across realistic scenarios and handles data of arbitrary dimensions by leveraging large-sample asymptotics. Through extensive simulations and a comparison against unsupervised tests of stratification based on random matrix theory, we find that our test compares favorably in various scenarios of interest. We apply the test to data from the 1000 Genomes Project, demonstrating how it can be employed to assess exchangeability of the genetic sample or find optimal linkage disequilibrium (LD) splits for downstream analysis. For exchangeability assessment we find that removing rare variants can substantially increase the p-value of the test statistic. For optimal LD splitting, the V test reports different optimal splits than previous approaches not relying on hypothesis testing. Software for our methods is available in R (CRAN: flintyR) and Python (PyPI: flintyPy).

REFERENCES

ANGELOPOULOS, A. N. and BATES, S. (2023). Conformal prediction: A gentle introduction. *Found. Trends Mach. Learn.* **16** 494–591.

AW, A. J., SPENCE, J. P. and SONG, Y. S. (2024). Supplement to "A simple and flexible test of sample exchangeability with applications to statistical genomics." https://doi.org/10.1214/23-AOAS1817SUPPA, https://doi.org/10.1214/23-AOAS1817SUPPB

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

BALASUBRAMANIAN, V., HO, S.-S. and VOVK, V. (2014). *Conformal Prediction for Reliable Machine Learning*: *Theory, Adaptations and Applications*. Morgan Kaufmann, San Mateo.

BARTELS, R. (1982). The rank version of von Neumann's ratio test for randomness. *J. Amer. Statist. Assoc.* **77** 40–46.

BERISA, T. and PICKRELL, J. K. (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32** 283–285. https://doi.org/10.1093/bioinformatics/btv546

BONNINI, S., CORAIN, L., MAROZZI, M. and SALMASO, L. (2014). *Nonparametric Hypothesis Testing*: *Rank and Permutation Methods with Applications in R*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3237553 https://doi.org/10.1002/9781118763490

DRAPER, D., HODGES, J. S., MALLOWS, C. L. and PREGIBON, D. (1993). Exchangeability and data analysis. *J. Roy. Statist. Soc. Ser. A* **156** 9–28. MR1207984 https://doi.org/10.2307/2982858

Key words and phrases. Exchangeability, feature independence, nonparametric test, population stratification, LD splitting.

Efron, B. (2009). Are a set of microarrays independent of each other? *Ann. Appl. Stat.* **3** 922–942. MR2750220 https://doi.org/10.1214/09-AOAS236

1000 Genomes Project Consortium et al. (2015). A global reference for human genetic variation. *Nature* **526** 68.

Gazal, S., Sahbatou, M., Babron, M.-C., Génin, E. and Leutenegger, A.-L. (2015). High level of inbreeding in final phase of 1000 Genomes Project. *Sci. Rep.* **5** 17453. https://doi.org/10.1038/srep17453

Gretton, A. and Györfi, L. (2010). Consistent nonparametric tests of independence. *J. Mach. Learn. Res.* **11** 1391–1423. MR2645456

Guo, L. and Modarres, R. (2020). Nonparametric tests of independence based on interpoint distances. *J. Nonparametr. Stat.* **32** 225–245. MR4077760 https://doi.org/10.1080/10485252.2020.1714613

Han, F., Chen, S. and Liu, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika* **104** 813–828. MR3737306 https://doi.org/10.1093/biomet/asx050

Heller, R. and Heller, Y. (2016). Multivariate tests of association based on univariate tests. *Adv. Neural Inf. Process. Syst.* **29**.

Hemerik, J. and Goeman, J. (2018). Exact testing with random permutations. *TEST* **27** 811–825. MR3878362 https://doi.org/10.1007/s11749-017-0571-1

Hernán, M. A. and Robins, J. M. (2020). *Causal Inference*: *What If*. Chapman & Hall/CRC Press, Boca Raton.

Kalina, J. and Janáček, P. (2022). Testing exchangeability of multivariate distributions. *J. Appl. Stat.* 1–15.

Kim, S. A., Cho, C.-S., Kim, S.-R., Bull, S. B. and Yoo, Y. J. (2018). A new haplotype block detection method for dense genome sequencing data based on interval graph modeling of clusters of highly correlated SNPs. *Bioinformatics* **34** 388–397.

Kingman, J. F. C. (1978). Uses of exchangeability. *Ann. Probab.* **6** 183–197. MR0494344 https://doi.org/10.1214/aop/1176995566

Kuchibhotla, A. K. (2020). Exchangeability, conformal prediction, and rank tests. Preprint. Available at arXiv:2005.06095.

Lindley, D. V. and Novick, M. R. (1981). The role of exchangeability in inference. *Ann. Statist.* **9** 45–58. MR0600531

Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X. and Sham, P. C. (2017). Polygenic scores via penalized regression on summary statistics. *Genet. Epidemiol.* **41** 469–480.

Mancuso, N., Freund, M. K., Johnson, R., Shi, H., Kichaev, G., Gusev, A. and Pasaniuc, B. (2019). Probabilistic fine-mapping of transcriptome-wide association studies. *Nat. Genet.* **51** 675–682. https://doi.org/10.1038/s41588-019-0367-1

Modarres, R. (2008). Tests of bivariate exchangeability. *Int. Stat. Rev.* **76** 203–213.

Mourad, R., Sinoquet, C., Dina, C. and Leray, P. (2011). Visualization of pairwise and multilocus linkage disequilibrium structure using latent forests. *PLoS ONE* **6** e27320. https://doi.org/10.1371/journal.pone.0027320

Patterson, N. et al. (2006). Population structure and eigenanalysis. *PLoS Genet.* **2** e190.

Pfister, N., Bühlmann, P., Schölkopf, B. and Peters, J. (2018). Kernel-based tests for joint independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 5–31. MR3744710 https://doi.org/10.1111/rssb.12235

Phipson, B. and Smyth, G. K. (2010). Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* **9** Art. 39, 14 pp. MR2746025 https://doi.org/10.2202/1544-6115.1585

Price, A. L., Weale, M. E., Patterson, N., Myers, S. R., Need, A. C., Shianna, K. V., Ge, D., Rotter, J. I., Torres, E. et al. (2008). Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83** 132–135.

Privé, F. (2022). Optimal linkage disequilibrium splitting. *Bioinformatics* **38** 255–256.

Privé, F., Arbel, J., Aschard, H. and Vilhjálmsson, B. (2021). Identifying and correcting for misspecifications in GWAS summary statistics and polygenic scores. *HGG Advances* **3** 100136.

Shafer, G. and Vovk, V. (2008). A tutorial on conformal prediction. *J. Mach. Learn. Res.* **9** 371–421. MR2417240

Shirts, B. H., Pritchard, C. C. and Walsh, T. (2016). Family-specific variants and the limits of human genetics. *Trends Mol. Med.* **22** 925–934. https://doi.org/10.1016/j.molmed.2016.09.007

Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108** 1033–1056. MR1933444 https://doi.org/10.1023/A:1019739414239

Spence, J. P., Sinnott-Armstrong, N., Assimes, T. and Pritchard, J. K. (2022). A flexible modeling and inference framework for estimating variant effect sizes from GWAS summary statistics. *bioRxiv*.

Spence, J. P. and Song, Y. S. (2019). Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Sci. Adv.* **5** eaaw9206.

TCHETGEN TCHETGEN, E. J., YING, A., CUI, Y., SHI, X. and MIAO, W. (2020). An introduction to proximal causal learning. Preprint. Available at arXiv:2009.10982.

TRACY, C. A. and WIDOM, H. (2002). Distribution functions for largest eigenvalues and their applications. In *Proceedings of the International Congress of Mathematicians*, *Vol. I* (*Beijing*, 2002) 587–596. Higher Ed. Press, Beijing. MR1989209

VOVK, V. (2021). Testing randomness online. *Statist. Sci.* **36** 595–611. MR4323055 https://doi.org/10.1214/20-sts817

WALD, A. and WOLFOWITZ, J. (1943). An exact test for randomness in the non-parametric case based on serial correlation. *Ann. Math. Stat.* **14** 378–388. MR0009838 https://doi.org/10.1214/aoms/1177731358

ZHOU, Y.-H., MARRON, J. S. and WRIGHT, F. A. (2018). Eigenvalue significance testing for genetic association. *Biometrics* **74** 439–447. MR3825330 https://doi.org/10.1111/biom.12767

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 882–898
https://doi.org/10.1214/23-AOAS1818

LEARNING HEALTHCARE DELIVERY NETWORK WITH LONGITUDINAL ELECTRONIC HEALTH RECORDS DATA

BY JIEHUAN SUN[1,a], KATHERINE P. LIAO[2,b] AND TIANXI CAI[3,c]

[1]*Division of Epidemiology and Biostatistics, University of Illinois at Chicago,* [a]*jiehuan@uic.edu*
[2]*Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital,* [b]*kliao@bwh.harvard.edu*
[3]*Department of Biostatistics, Harvard T.H. Chan School of Public Health,* [c]*tcai@hsph.harvard.edu*

Knowledge networks, such as the healthcare delivery network (HDN), describing relationships among different medical encounters, are useful summaries of state-of-art medical knowledge. The increasing availability of longitudinal electronic health records (EHR) data promises a rich data source for learning HDN. Most existing methods for inferring knowledge networks are based on cooccurrence patterns that do not account for temporal effects or patient-level heterogeneity. In this article, building upon the multivariate Hawkes process (mvHP), we propose a flexible covariate-adjusted random effects (CARE) mvHP modeling strategy for HDN construction. Our model allows for patient-specific time-varying background intensity functions via random effects, which can also adjust for effects of important covariates. We adopt a penalized approach to select fixed effects, yielding a sparse network structure, and to remove unnecessary random effects from the model. Through extensive simulation studies, we show that our proposed method performs well in recovering the network structure and that it is essential to account for patient heterogeneities. We further illustrate our CARE mvHP method in an EHR study of type 2 diabetes patients to learn an HDN for these patients and demonstrate that our results are consistent with current clinical practice in healthcare systems.

REFERENCES

BACRY, E., GAÏFFAS, S. and MUZY, J.-F. (2015). A generalization error bound for sparse and low-rank multivariate Hawkes processes. ArXiv Preprint. Available at arXiv:1501.00725.

BATES, D. M. and DEBROY, S. (2004). Linear mixed models and penalized least squares. *J. Multivariate Anal.* **91** 1–17. MR2083901 https://doi.org/10.1016/j.jmva.2004.04.013

BERRY, T., HAMILTON, F., PEIXOTO, N. and SAUER, T. (2012). Detecting connectivity changes in neuronal networks. *J. Neurosci. Methods* **209** 388–397. https://doi.org/10.1016/j.jneumeth.2012.06.021

BONDELL, H. D., KRISHNA, A. and GHOSH, S. K. (2010). Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics* **66** 1069–1077. MR2758494 https://doi.org/10.1111/j.1541-0420.2010.01391.x

BRÉMAUD, P. and MASSOULIÉ, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Probab.* **24** 1563–1588. MR1411506 https://doi.org/10.1214/aop/1065725193

BRILLINGER, D. R. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol. Cybernet.* **59** 189–200. https://doi.org/10.1007/BF00318010

BROWN, E. N., KASS, R. E. and MITRA, P. P. (2004). Multiple neural spike train data analysis: State-of-the-art and future challenges. *Nat. Neurosci.* **7** 456–461.

BURNHAM, K. P. and WHITE, G. C. (2002). Evaluation of some random effects methodology applicable to bird ringing data. *J. Appl. Stat.* **29** 245–264.

CARSTENSEN, L., SANDELIN, A., WINTHER, O. and HANSEN, N. R. (2010). Multivariate Hawkes process models of the occurrence of regulatory elements. *BMC Bioinform.* **11** 456.

CHEN, S., SHOJAIE, A., SHEA-BROWN, E. and WITTEN, D. (2017). The multivariate Hawkes process in high dimensions: Beyond mutual excitation. ArXiv Preprint. Available at arXiv:1707.04928.

Key words and phrases. Electronic health records, group lasso, Hawkes process, healthcare delivery network, random effects selection.

CHOI, E., DU, N., CHEN, R., SONG, L. and SUN, J. (2015). Constructing disease network and temporal progression model via context-sensitive Hawkes process. In *IEEE International Conference on Data Mining* (*ICDM*) 721–726.

COX, D. R. and ISHAM, V. (1980). *Point Processes*. *Monographs on Applied Probability and Statistics*. CRC Press, London. MR0598033

DENNY, J. C., BASTARACHE, L., RITCHIE, M. D., CARROLL, R. J., ZINK, R., MOSLEY, J. D., FIELD, J. R., PULLEY, J. M., RAMIREZ, A. H. et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat*. *Biotechnol*. **31** 1102–1111.

DIGGLE, P. J., MORAGA, P., ROWLINGSON, B. and TAYLOR, B. M. (2013). Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm. *Statist*. *Sci*. **28** 542–563. MR3161587 https://doi.org/10.1214/13-STS441

EICHLER, M., DAHLHAUS, R. and DUECK, J. (2017). Graphical modeling for multivariate Hawkes processes with nonparametric link functions. *J*. *Time Series Anal*. **38** 225–242. MR3611742 https://doi.org/10.1111/jtsa.12213

EMBRECHTS, P., LINIGER, T. and LIN, L. (2011). Multivariate Hawkes processes: An application to financial data. *J*. *Appl*. *Probab*. **48A** 367–378. MR2865638 https://doi.org/10.1239/jap/1318940477

GOMEZ-RODRIGUEZ, M., LESKOVEC, J. and KRAUSE, A. (2012). Inferring networks of diffusion and influence. *ACM Trans*. *Knowl*. *Discov*. *Data* **5** 1–37.

HANSEN, N. R., REYNAUD-BOURET, P. and RIVOIRARD, V. (2015). Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli* **21** 83–143. MR3322314 https://doi.org/10.3150/13-BEJ562

HARTE, D. et al. (2010). PtProcess: An R package for modelling marked point processes indexed by time. *J*. *Stat*. *Softw*. **35** 1–32.

HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90. MR0278410 https://doi.org/10.1093/biomet/58.1.83

HIDALGO, C. A., BLUMM, N., BARABÁSI, A.-L. and CHRISTAKIS, N. A. (2009). A dynamic network approach for the study of human phenotypes. *PLoS Comput*. *Biol*. **5** e1000353.

JIANG, Y., MA, S., SHIA, B.-C. and LEE, T.-S. (2018). An epidemiological human disease network derived from disease co-occurrence in Taiwan. *Sci*. *Rep*. **8** 1–12.

LIAO, K. P., CAI, T., SAVOVA, G. K., MURPHY, S. N., KARLSON, E. W., ANANTHAKRISHNAN, A. N., GAINER, V. S., SHAW, S. Y., XIA, Z. et al. (2015). Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *BMJ* **350** h1885.

LUO, D., XU, H., ZHEN, Y., NING, X., ZHA, H., YANG, X. and ZHANG, W. (2015). Multi-task multi-dimensional Hawkes processes for modeling event sequences. In *Proceedings of the* 24*th International Conference on Artificial Intelligence* 3685–3691. AAAI Press, Menlo Park.

MASUD, M. S. and BORISYUK, R. (2011). Statistical technique for analysing functional connectivity of multiple spike trains. *J*. *Neurosci*. *Methods* **196** 201–219. https://doi.org/10.1016/j.jneumeth.2011.01.003

MØLLER, J., SYVERSVEEN, A. R. and WAAGEPETERSEN, R. P. (1998). Log Gaussian Cox processes. *Scand*. *J*. *Stat*. **25** 451–482. MR1650019 https://doi.org/10.1111/1467-9469.00115

OGATA, Y. (1981). On Lewis' simulation method for point processes. *IEEE Trans*. *Inf*. *Theory* **27** 23–31.

OKATAN, M., WILSON, M. A. and BROWN, E. N. (2005). Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural Comput*. **17** 1927–1961.

PEROTTE, A., RANGANATH, R., HIRSCH, J. S., BLEI, D. and ELHADAD, N. (2015). Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis. *J*. *Amer*. *Med*. *Inform*. *Assoc*. **22** 872–880.

REYNAUD-BOURET, P. and SCHBATH, S. (2010). Adaptive estimation for Hawkes processes; application to genome analysis. *Ann*. *Statist*. **38** 2781–2822. MR2722456 https://doi.org/10.1214/10-AOS806

ROTMENSCH, M., HALPERN, Y., TLIMAT, A., HORNG, S. and SONTAG, D. (2017). Learning a health knowledge graph from electronic medical records. *Sci*. *Rep*. **7** 5994. https://doi.org/10.1038/s41598-017-05778-z

RZHETSKY, A., WAJNGURT, D., PARK, N. and ZHENG, T. (2007). Probing genetic overlap among complex human phenotypes. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **104** 11694–11699.

SCHELLDORFER, J., MEIER, L. and BÜHLMANN, P. (2014). GLMMLasso: An algorithm for high-dimensional generalized linear mixed models using ℓ_1-penalization. *J*. *Comput*. *Graph*. *Statist*. **23** 460–477. MR3215820 https://doi.org/10.1080/10618600.2013.773239

SUN, J., LIAO, P. K. and CAI, T. (2024). Supplement to "Learning healthcare delivery network with longitudinal electronic health records data." https://doi.org/10.1214/23-AOAS1818SUPP

VAIDA, F. and BLANCHARD, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika* **92** 351–370. MR2201364 https://doi.org/10.1093/biomet/92.2.351

WAINWRIGHT, M. J. and JORDAN, M. I. (2008). *Graphical Models*, *Exponential Families*, *and Variational Inference*. Now Publishers, Hanover.

WANG, S., SONG, P. and ZHU, J. (2010). Doubly regularized REML for estimation and selection of fixed and random effects in linear mixed-effects models. The University of Michigan Department of Biostatistics Working Paper Series.

XU, G., WANG, M., BIAN, J., HUANG, H., BURCH, T. R., ANDRADE, S. C., ZHANG, J. and GUAN, Y. (2020). Semi-parametric learning of structured temporal point processes. *J. Mach. Learn. Res.* **21** 192. MR4209478

XU, H., FARAJTABAR, M. and ZHA, H. (2016). Learning granger causality for Hawkes processes. In *Proceedings of the* 33*rd International Conference on International Conference on Machine Learning-Volume* 48 1660–1669. JMLR.org.

YU, S., LIAO, K. P., SHAW, S. Y., GAINER, V. S., CHURCHILL, S. E., SZOLOVITS, P., MURPHY, S. N., KOHANE, I. S. and CAI, T. (2015). Toward high-throughput phenotyping: Unbiased automated feature extraction and selection from knowledge sources. *J. Amer. Med. Inform. Assoc.* **22** 993–1000.

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

ZHOU, K., ZHA, H. and SONG, L. (2013). Learning social infectivity in sparse low-rank networks using multidimensional Hawkes processes. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics* (C. M. Carvalho and P. Ravikumar, eds.). *Proceedings of Machine Learning Research* **31** 641–649. PMLR, Scottsdale, AZ, USA.

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

The Annals of Applied Statistics
2024, Vol. 18, No. 1, 899–903
https://doi.org/10.1214/23-AOAS1861
Main article: https://doi.org/10.1214/22-AOAS1642

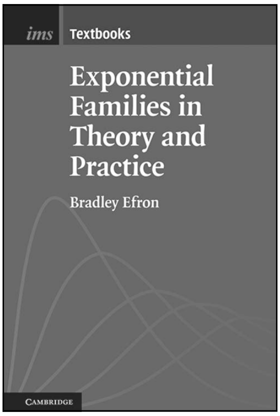
CORRECTION TO: SPATIOTEMPORAL WILDFIRE MODELING THROUGH POINT PROCESSES WITH MODERATE AND EXTREME MARKS

BY JONATHAN KOH[1,2,a], FRANÇOIS PIMONT[3,b], JEAN-LUC DUPUY[3,c] AND THOMAS OPITZ[4,d]

[1]*Institute of Mathematics, EPFL*
[2]*Institute of Mathematical Statistics and Actuarial Science, Oeschger Centre for Climate Change Research, University of Bern,* [a]*jonathan.koh@unibe.ch*
[3]*URFM UR629, INRAE,* [b]*francois.pimont@inrae.fr,* [c]*jean-luc.dupuy@inrae.fr*
[4]*BioSP UR546, INRAE,* [d]*thomas.opitz@inrae.fr*

REFERENCES

KOH, J., PIMONT, F., DUPUY, J.-L. and OPITZ, T. (2023). Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *Ann. Appl. Stat.* **17** 560–582. MR4539044 https://doi.org/10.1214/22-aoas1642



NEW BOOK

The Institute of Mathematical Statistics presents

IMS TEXTBOOKS

Exponential Families in Theory and Practice

Bradley Efron, Stanford University

During the past half-century, exponential families have attained a position at the center of parametric statistical inference. Theoretical advances have been matched, and more than matched, in the world of applications, where logistic regression by itself has become the go-to methodology in medical statistics, computer-based prediction algorithms, and the social sciences. This book is based on a one-semester graduate course for first year Ph.D. and advanced master's students. After presenting the basic structure of univariate and multivariate exponential families, their application to generalized linear models including logistic and Poisson regression is described in detail, emphasizing geometrical ideas, computational practice, and the analogy with ordinary linear regression. Connections are made with a variety of current statistical methodologies: missing data, survival analysis and proportional hazards, false discovery rates, bootstrapping, and empirical Bayes analysis. The book connects exponential family theory with its applications in a way that doesn't require advanced mathematical preparation.

Hardback $105.00

Paperback $39.99

IMS members are entitled to a 40% discount: email ims@imstat.org to request your code

www.imstat.org/cup/

Cambridge University Press, with the Institute of Mathematical Statistics, established the *IMS Monographs* and *IMS Textbooks* series of high-quality books. The series editors are Mark Handcock, Ramon van Handel, Arnaud Doucet, and John Aston.