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

THE ANNALS *of* APPLIED STATISTICS

An Official Journal of the
Institute of Mathematical Statistics

Articles

Causal Inference, Design and Policy

Computational Biology

Economics and Finance

Environmental Science

Machine Learning and Artificial Intelligence

Medical and Health Sciences

continued

Vol. 20, No. 1—March 2026

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

THE ANNALS *of* APPLIED STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Continued from back cover

Physical Science and Engineering

Prospective Paper

Reinforcement Learning

Social and Political Sciences

Sports

Vol. 20, No. 1—March 2026

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

THE ANNALS *of* APPLIED STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

continued

Continued on inner cover

Vol. 20, No. 1—March 2026

THE ANNALS OF APPLIED STATISTICS Vol. 20, No. 1, pp. 1–892 March 2026

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

***The Annals of Applied Statistics* [ISSN 1932-6157 (print); ISSN 1941-7330 (online)],** Volume 20, Number 1, March 2026. 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.

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 1–24
https://doi.org/10.1214/25-AOAS2117

MULTIPLY ROBUST ESTIMATION FOR CAUSAL SURVIVAL ANALYSIS WITH TREATMENT NONCOMPLIANCE

BY CHAO CHENG[1,a], BO LIU[2,b], LISA WRUCK[3,4,d], FAN LI[2,c] AND FAN LI[5,e]

[1]*Department of Statistics and Data Science, Washington University in St. Louis,* [a]*chaoc@wustl.edu*
[2]*Department of Statistical Science, Duke University,* [b]*bo.liu1997@duke.edu,* [c]*fl35@duke.edu*
[3]*Department of Biostatistics and Bioinformatics, Duke University,* [d]*lisa.wruck@duke.edu*
[4]*Duke Clinical Research Institute, Duke University*
[5]*Department of Biostatistics, Yale University,* [e]*fan.f.li@yale.edu*

Comparative effectiveness research frequently addresses a time-to-event outcome and can require unique considerations in the presence of treatment noncompliance. Motivated by the challenges in addressing noncompliance in the ADAPTABLE pragmatic clinical trial, we develop a multiply robust estimator to estimate the principal survival causal effects under the principal ignorability and monotonicity. The multiply robust estimator is consistent, even if one, and sometimes two, of the required models are misspecified. We apply the multiply robust method in the ADAPTABLE trial to evaluate the effect of low- vs. high-dose aspirin assignment on patients' death and hospitalization from cardiovascular diseases. We find that, comparing to low-dose assignment, assignment to the high-dose leads to differential effects among always high-dose takers, compliers, and always low-dose takers. Such treatment effect heterogeneity contributes to the null intention-to-treatment effect. We further perform a formal sensitivity analysis for investigating the robustness of our causal conclusions under violation of two identification assumptions specific to noncompliance.

REFERENCES

ANGELUCCI, M., ATTANASIO, O. and DI MARO, V. (2012). The impact of Oportunidades on consumption, savings and transfers. *Fisc. Stud.* **33** 305–334.

ANGELUCCI, M. and ATTANASIO, O. P. (2006). Estimating ATT effects with non-experimental data and low compliance. Technical Report, IZA Discussion Papers.

ANGRIST, J., IMBENS, G. and RUBIN, D. (1996). Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* **91** 444–455.

BAI, X., TSIATIS, A. A. and O'BRIEN, S. M. (2013). Doubly-robust estimators of treatment-specific survival distributions in observational studies with stratified sampling. *Biometrics* **69** 830–839. MR3146779 https://doi.org/10.1111/biom.12076

BAKER, S. G. (1998). Analysis of survival data from a randomized trial with all-or-none compliance: Estimating the cost-effectiveness of a cancer screening program. *J. Amer. Statist. Assoc.* **93** 929–934.

BAKER, S. G. and LINDEMAN, K. S. (1994). The paired availability design: A proposal for evaluating epidural analgesia during labor. *Stat. Med.* **13** 2269–2278.

BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–973. MR2216189 https://doi.org/10.1111/j.1541-0420.2005.00377.x

BRESLOW, N. E. (1972). Discussion of professor Cox's paper. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **34** 216–217.

CHENG, C., GUO, Y., LIU, B., WRUCK, L. and LI, F. (2023). Multiply robust estimation for causal survival analysis with treatment noncompliance. arXiv preprint. Available at arXiv:2305.13443v2.

CHENG, C., LIU, B., WRUCK, L., LI, F. and LI, F. (2026). Supplement to "Multiply robust estimation for causal survival analysis with treatment noncompliance." https://doi.org/10.1214/25-AOAS2117SUPPA, https://doi.org/10.1214/25-AOAS2117SUPPB

CINELLI, C. and HAZLETT, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 39–67. MR4060976

Key words and phrases. Efficient influence function, pragmatic clinical trials, principal stratification, principal score, sensitivity analysis, time-to-event data.

CUZICK, J., SASIENI, P., MYLES, J. and TYRER, J. (2007). Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 565–588. MR2370069 https://doi.org/10.1111/j.1467-9868.2007.00600.x

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

FRANGAKIS, C. E. and RUBIN, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* **86** 365–379. MR1705410 https://doi.org/10.1093/biomet/86.2.365

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

FRANK, K. A. (2000). Impact of a confounding variable on a regression coefficient. *Sociol. Methods Res.* **29** 147–194.

GILL, R. D., LAAN, M. J. and ROBINS, J. M. (1997). Coarsening at random: Characterizations, conjectures, counter-examples. In *Proceedings of the First Seattle Symposium in Biostatistics* 255–294 Springer, Berlin.

HIRANO, K., IMBENS, G., RUBIN, D. and ZHOU, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* **1** 69–88.

IMBENS, G. and ANGRIST, J. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62** 467–475.

JIANG, Z., YANG, S. and DING, P. (2022). Multiply robust estimation of causal effects under principal ignorability. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1423–1445. MR4494165

JO, B. and STUART, E. A. (2009). On the use of propensity scores in principal causal effect estimation. *Stat. Med.* **28** 2857–2875. MR2750169 https://doi.org/10.1002/sim.3669

JONES, W. S., MULDER, H., WRUCK, L. M., PENCINA, M. J., KRIPALANI, S., MUÑOZ, D., CRENSHAW, D. L., EFFRON, M. B., RE, R. N. et al. (2021). Comparative effectiveness of aspirin dosing in cardiovascular disease. *N. Engl. J. Med.* **384** 1981–1990.

KAHAN, B. C., CRO, S., LI, F. and HARHAY, M. O. (2023). Eliminating ambiguous treatment effects using estimands. *Amer. J. Epidemiol.* **192** 987–994.

LI, F., BUCHANAN, A. L. and COLE, S. R. (2022). Generalizing trial evidence to target populations in non-nested designs: Applications to AIDS clinical trials. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **71** 669–697. MR4441621 https://doi.org/10.1111/rssc.12550

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

LI, F., MORGAN, K. L. and ZASLAVSKY, A. M. (2018). Balancing covariates via propensity score weighting. *J. Amer. Statist. Assoc.* **113** 390–400. MR3803473 https://doi.org/10.1080/01621459.2016.1260466

LIU, B., WRUCK, L. and LI, F. (2024). Principal stratification analysis of noncompliance with time-to-event outcomes. *Biometrics* **80** Paper No. ujad016, 14. MR4856610 https://doi.org/10.1093/biomtc/ujad016

LOEYS, T. and GOETGHEBEUR, E. (2003). A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. *Biometrics* **59** 100–105. MR1978476 https://doi.org/10.1111/1541-0420.00012

MCCANDLESS, L. C., GUSTAFSON, P. and LEVY, A. (2007). Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Stat. Med.* **26** 2331–2347. MR2368419 https://doi.org/10.1002/sim.2711

NGUYEN, T. Q., CARLSON, M. C. and STUART, E. A. (2024). Identification of complier and noncomplier average causal effects in the presence of *latent* missing-at-random (LMAR) outcomes: A unifying view and choices of assumptions. *Biostatistics* **25** 978–996. MR4808868 https://doi.org/10.1093/biostatistics/kxae011

NIE, H., CHENG, J. and SMALL, D. S. (2011). Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring. *Biometrics* **67** 1397–1405. MR2872390 https://doi.org/10.1111/j.1541-0420.2011.01575.x

ROBINS, J. M. and FINKELSTEIN, D. M. (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* **56** 779–788.

STÜRMER, T., GLYNN, R. J., ROTHMAN, K. J., AVORN, J. and SCHNEEWEISS, S. (2007). Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. *Med. Care* **45** S158–S165.

TONG, J., KAHAN, B., HARHAY, M. O. and LI, F. (2025). Semiparametric principal stratification analysis beyond monotonicity. arXiv preprint. Available at arXiv:2501.17514.

TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data. Springer Series in Statistics*. Springer, New York. MR2233926

VANDERWEELE, T. J. and DING, P. (2017). Sensitivity analysis in observational research: Introducing the E-value. *Ann. Intern. Med.* **167** 268–274.

WEI, B., PENG, L., ZHANG, M.-J. and FINE, J. P. (2021). Estimation of causal quantile effects with a binary instrumental variable and censored data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 559–578. MR4294544 https://doi.org/10.1111/rssb.12431

WESTLING, T., LUEDTKE, A., GILBERT, P. B. and CARONE, M. (2024). Inference for treatment-specific survival curves using machine learning. *J. Amer. Statist. Assoc.* **119** 1541–1553. MR4766008 https://doi.org/10.1080/01621459.2023.2205060

YU, W., CHEN, K., SOBEL, M. E. and YING, Z. (2015). Semiparametric transformation models for causal inference in time-to-event studies with all-or-nothing compliance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 397–415. MR3310532 https://doi.org/10.1111/rssb.12072

ZENG, S., LI, F., WANG, R. and LI, F. (2021). Propensity score weighting for covariate adjustment in randomized clinical trials. *Stat. Med.* **40** 842–858. MR4201106 https://doi.org/10.1002/sim.8805

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 25–45
https://doi.org/10.1214/25-AOAS2118

INFERRING THE EFFECT OF A RANDOMISED TREATMENT ON A RECURRENT EVENT PROCESS UNDER DEPENDENT CENSORING

BY WOUT WATERSCHOOT[1,a], ANDREA CALLEGARO[2,c], LUCA MORASCHINI[3,d] AND STIJN VANSTEELANDT[1,b]

[1]*Department of Mathematics, Computer Science and Statistics, Ghent University,* [a]*wout.waterschoot@ugent.be,* [b]*stijn.vansteelandt@ugent.be*

[2]*Department of Biostatistics, GlaxoSmithKline,* [c]*andrea.x.callegaro@gsk.com*

[3]*Department of Vaccines Clinical Statistics, GlaxoSmithKline,* [d]*luca.x.moraschini@gsk.com*

This work is motivated by randomized clinical trial NCT03281876 (November 2017–March 2020), whose secondary aim was to evaluate the efficacy of the NTHi-Mcat vaccine vs. placebo in preventing recurrent severe exacerbations among patients with acute exacerbations of chronic obstructive pulmonary disease (AECOPD). The published analysis (*Vaccine* **40** (2022) 5924–5932; *Lancet Respir. Med.* **10** (2022) 435–446) aimed to estimate the ratio of the expected number of exacerbations one experienced by the end of study in the vaccinated vs. the placebo arm. One, therefore, regressed the number of exacerbations one experienced by the last point in time one is uncensored on treatment and baseline covariates with offset the logarithm of the observation time to account for different follow-up times. In this paper we demonstrate that this approach is prone to selection bias due to: (i) selective withdrawal and (ii) selective timing of the outcome measurements. We show that inverse probability of censoring weigting (IPCW), a common approach to adjust for dependent censoring under the assumption that censoring is non-informative given the observed covariate history, does not suffice to restore the unbiasedness of the treatment effect estimator under the above-mentioned type of analysis. To address this, we propose hazard inverse probability of censoring weighting (HIPCW). This novel weighting technique preserves the simplicity of IPCW but extracts more efficiency by using each individual's last recorded outcome. We validate the proposed approach through extensive simulations and compare with: (i) IPCW at a single time, (ii) a variant of the existing IPCW-GEE routines (*J. Amer. Statist. Assoc.* **90** (1995) 106–121) and (iii) IPCW-based estimators derived from the popular Andersen-Gill model (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** (2004) 239–257). We illustrate the routines through reanalysing clinical trial NCT03281876.

REFERENCES

AMORIM, L. D. and CAI, J. (2015). Modelling recurrent events: A tutorial for analysis in epidemiology. *Int. J. Epidemiol.* **44** 324–333. https://doi.org/10.1093/ije/dyu222

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. *Springer Series in Statistics*. Springer, New York. MR1198884 https://doi.org/10.1007/978-1-4612-4348-9

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. MR0673646

ANDREAS, S., TESTA, M., BOYER, L., BRUSSELLE, G., JANSSENS, W., KERWIN, E., PAPI, A., PEK, B., PUENTE-MAESTU, L. et al. (2022). Non-typeable Haemophilus influenzae–Moraxella catarrhalis vaccine for the prevention of exacerbations in chronic obstructive pulmonary disease: A multicentre, randomised, placebo-controlled, observer-blinded, proof-of-concept, phase 2b trial. *Lancet Respir. Med.* **10** 435–446. https://doi.org/10.1016/S2213-2600(21)00502-6

Key words and phrases. Recurrent event processes, dependent censoring, inverse probability of censoring weighting (IPCW), vaccine efficacy.

Arora, A. K., Chinsky, K., Keller, C., Mayers, I., Pascual-Guardia, S., Vera, M. P., Lambert, C., Lombardi, S., Rondini, S. et al. (2022). A detailed analysis of possible efficacy signals of NTHi-Mcat vaccine against severe COPD exacerbations in a previously reported randomised phase 2b trial. *Vaccine* **40** 5924–5932. https://doi.org/10.1016/j.vaccine.2022.08.053

Bacharier, L. B., Maspero, J. F., Katelaris, C. H., Fiocchi, A. G., Gagnon, R., De Mir, I., Jain, N., Sher, L. D., Mao, X. et al. (2021). Dupilumab in children with uncontrolled moderate-to-severe asthma. *N. Engl. J. Med.* **385** 2230–2240. https://doi.org/10.1056/NEJMoa2106567

Blanche, P. F., Holt, A. and Scheike, T. (2023). On logistic regression with right censored data, with or without competing risks, and its use for estimating treatment effects. *Lifetime Data Anal.* **29** 441–482. MR4559048 https://doi.org/10.1007/s10985-022-09564-6

Boos, D. D. and Stefanski, L. A. (2013). *Essential Statistical Inference*: *Theory and Methods*. *Springer Texts in Statistics*. Springer, New York. MR3024617 https://doi.org/10.1007/978-1-4614-4818-1

Cai, J. and Schaubel, D. E. (2004). Analysis of recurrent event data. In *Advances in Survival Analysis*. *Handbook of Statist.* **23** 603–623. Elsevier, Amsterdam. MR2065791 https://doi.org/10.1016/S0169-7161(03)23034-0

Cameron, A. C. and Trivedi, P. K. (2013). *Regression Analysis of Count Data*, 2nd ed. *Econometric Society Monographs* **53**. Cambridge Univ. Press, Cambridge. MR3155491 https://doi.org/10.1017/CBO9781139013567

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. *Statistics for Biology and Health*. Springer, New York. MR3822124

Cook, R. J., Lawless, J. F., Lakhal-Chaieb, L. and Lee, K.-A. (2009). Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination: Application to skeletal complications in cancer metastatic to bone. *J. Amer. Statist. Assoc.* **104** 60–75. MR2502850 https://doi.org/10.1198/jasa.2009.0004

Cortese, G. and Scheike, T. H. (2022). Efficient estimation of the marginal mean of recurrent events. *J. R. Stat. Soc.*, *Ser. C*, *Appl. Stat.* **71** 1787–1821. MR4511131 https://doi.org/10.1111/rssc.12586

Ghosh, D. and Lin, D. Y. (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* **59** 877–885. MR2025111 https://doi.org/10.1111/j.0006-341X.2003.00102.x

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

Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology* **21** 13–15. https://doi.org/10.1097/EDE.0b013e3181c1ea43

Hernán, M. A. (2016). Does water kill? A call for less casual causal inferences. *Ann. Epidemiol.* **26** 674–680. https://doi.org/10.1016/j.annepidem.2016.08.016

Hernán, M. A. and Robins, J. M. (2020). *Causal Inference*: *What If*. CRC Press/CRC, Boca Raton.

Hernán, M. A. and Taubman, S. L. (2008). Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int. J. Obes.* **32 Suppl 3** S8–14. https://doi.org/10.1038/ijo.2008.82

Hsu, L., Gorfine, M. and Malone, K. (2007). On robustness of marginal regression coefficient estimates and hazard functions in multivariate survival analysis of family data when the frailty distribution is mis-specified. *Stat. Med.* **26** 4657–4678. MR2411893 https://doi.org/10.1002/sim.2870

Huang, C.-Y. and Wang, M.-C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *J. Amer. Statist. Assoc.* **99** 1153–1165. MR2109503 https://doi.org/10.1198/016214504000001033

Lancaster, T. and Intrator, O. (1998). Panel data with survival: Hospitalization of HIV-positive patients. *J. Amer. Statist. Assoc.* **93** 46–53. https://doi.org/10.2307/2669601

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

Liu, L., Wolfe, R. A. and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60** 747–756. MR2089451 https://doi.org/10.1111/j.0006-341X.2004.00225.x

Lunceford, J. K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Stat. Med.* **23** 2937–2960. https://doi.org/10.1002/sim.1903

Mazroui, Y., Mathoulin-Pelissier, S., Soubeyran, P. and Rondeau, V. (2012). General joint frailty model for recurrent event data with a dependent terminal event: Application to follicular lymphoma data. *Stat. Med.* **31** 1162–1176. MR2925687 https://doi.org/10.1002/sim.4479

Miloslavsky, M., Keleş, S., van der Laan, M. J. and Butler, S. (2004). Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 239–257. MR2035769 https://doi.org/10.1111/j.1467-9868.2004.00442.x

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Model.* **7** 1393–1512. MR0877758 https://doi.org/10.1016/0270-0255(86)90088-6

ROBINS, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proc. Biopharm. Sect.* **24** 24–33.

ROBINS, J. M. (1997). Marginal structural models. In 1997 *Proceedings of the Section on Bayesian Statistical Science* 1–10. American Statistical Association, Alexandria, VA.

ROBINS, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology*, *the Environment*, *and Clinical Trials* (*Minneapolis*, *MN*, 1997). *IMA Vol. Math. Appl.* **116** 95–133. Springer, New York. MR1731682 https://doi.org/10.1007/978-1-4612-1284-3_2

ROBINS, J. M. and FINKELSTEIN, D. M. (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* **56** 779–788. https://doi.org/10.1111/j.0006-341x.2000.00779.x

ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.* **90** 106–121. MR1325118

ROGERS, J. K., YAROSHINSKY, A., POCOCK, S. J., STOKAR, D. and POGODA, J. (2016). Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. *Stat. Med.* **35** 2195–2205. MR3513508 https://doi.org/10.1002/sim.6853

RONDEAU, V., GONZALEZ, J. R., MAZROUI, Y., MAUGUEN, A., DIAKITE, A., LAURENT, A., LOPEZ, M., KROL, A., SOFEU, C. L. et al. (2021). frailtypack: Shared, Joint (Generalized) Frailty Models; Surrogate Endpoints.

RONDEAU, V., MATHOULIN-PELISSIER, S., JACQMIN-GADDA, H., BROUSTE, V. and SOUBEYRAN, P. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: Application on cancer events. *Biostatistics* **8** 708–721. https://doi.org/10.1093/biostatistics/kxl043

ROSENBAUM, P. R. (1984). The consquences of adjustment for a concomitant variable that has been affected by the treatment. *J. R. Stat. Soc. A*, *General* **147** 656–666. https://doi.org/10.2307/2981697

ROTNITZKY, A., ROBINS, J. M. and SCHARFSTEIN, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *J. Amer. Statist. Assoc.* **93** 1321–1339. MR1666631 https://doi.org/10.2307/2670049

SHU, D., YOUNG, J. G., TOH, S. and WANG, R. (2021). Variance estimation in inverse probability weighted Cox models. *Biometrics* **77** 1101–1117. MR4320681 https://doi.org/10.1111/biom.13332

STEINMAN, L., FOX, E., HARTUNG, H.-P., ALVAREZ, E., QIAN, P., WRAY, S., ROBERTSON, D., HUANG, D., SELMAJ, K. et al. (2022). Ublituximab vs. Teriflunomide in relapsing multiple sclerosis. *N. Engl. J. Med.* **387** 704–714. https://doi.org/10.1056/NEJMoa2201904

SU, L., SEAMAN, S. R. and YIU, S. (2022). Sensitivity analysis for calibrated inverse probability-of-censoring weighted estimators under non-ignorable dropout. *Stat. Methods Med. Res.* **31** 1374–1391. MR4446584 https://doi.org/10.1177/09622802221090763

TCHETGEN TCHETGEN, E. J., GLYMOUR, M. M., WEUVE, J. and ROBINS, J. (2012). Specifying the correlation structure in inverse-probability- weighting estimation for repeated measures. *Epidemiology* **23** 644–646. https://doi.org/10.1097/EDE.0b013e31825727b5

THERNEAU, T. M. (2023). A package for survival analysis in R.

THERNEAU, T. M. and GRAMBSCH, P. M. (2000). *Modeling Survival Data*: *Extending the Cox Model*. *Statistics for Biology and Health*. Springer, New York. MR1774977 https://doi.org/10.1007/978-1-4757-3294-8

TSIATIS, A. A., DAVIDIAN, M. and CAO, W. (2011). Improved doubly robust estimation when data are monotonely coarsened, with application to longitudinal studies with dropout. *Biometrics* **67** 536–545. MR2829022 https://doi.org/10.1111/j.1541-0420.2010.01476.x

VANDERWEELE, T. J. and HERNAN, M. A. (2013). Causal inference under multiple versions of treatment. *J. Causal Inference* **1** 1–20. MR4289399 https://doi.org/10.1515/jci-2012-0002

WATERSCHOOT, W., CALLEGARO, A., MORASCHINI, L. and VANSTEELANDT, S. (2026a). Supplement S1 to "Inferring the effect of a randomised treatment on a recurrent event process under dependent censoring." https://doi.org/10.1214/25-AOAS2118SUPPA

WATERSCHOOT, W., CALLEGARO, A., MORASCHINI, L. and VANSTEELANDT, S. (2026b). Supplement S2 to "Inferring the effect of a randomised treatment on a recurrent event process under dependent censoring." https://doi.org/10.1214/25-AOAS2118SUPPB

WATERSCHOOT, W., CALLEGARO, A., MORASCHINI, L. and VANSTEELANDT, S. (2026c). Supplement S3 to "Inferring the effect of a randomised treatment on a recurrent event process under dependent censoring." https://doi.org/10.1214/25-AOAS2118SUPPC

WATERSCHOOT, W., CALLEGARO, A., MORASCHINI, L. and VANSTEELANDT, S. (2026d). Supplement S4 to "Inferring the effect of a randomised treatment on a recurrent event process under dependent censoring." https://doi.org/10.1214/25-AOAS2118SUPPD

WATERSCHOOT, W., CALLEGARO, A., MORASCHINI, L. and VANSTEELANDT, S. (2026e). Supplement S5 to "Inferring the effect of a randomised treatment on a recurrent event process under dependent censoring." https://doi.org/10.1214/25-AOAS2118SUPPE

WATERSCHOOT, W., CALLEGARO, A., MORASCHINI, L. and VANSTEELANDT, S. (2026f). Supplement S6 to "Inferring the effect of a randomised treatment on a recurrent event process under dependent censoring." https://doi.org/10.1214/25-AOAS2118SUPPF

WATERSCHOOT, W., CALLEGARO, A., MORASCHINI, L. and VANSTEELANDT, S. (2026g). Supplement S7 to "Inferring the effect of a randomised treatment on a recurrent event process under dependent censoring." https://doi.org/10.1214/25-AOAS2118SUPPG

XU, G., CHIOU, S. H., HUANG, C.-Y., WANG, M.-C. and YAN, J. (2017). Joint scale-change models for recurrent events and failure time. *J. Amer. Statist. Assoc.* **112** 794–805. MR3671771 https://doi.org/10.1080/01621459.2016.1173557

YOUNG, J. G., STENSRUD, M. J., TCHETGEN TCHETGEN, E. J. and HERNÁN, M. A. (2020). A causal framework for classical statistical estimands in failure-time settings with competing events. *Stat. Med.* **39** 1199–1236. MR4075855 https://doi.org/10.1002/sim.8471

ZEGER, S. L., LIANG, K.-Y. and ALBERT, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44** 1049–1060. MR0980999 https://doi.org/10.2307/2531734

ZENG, D., IBRAHIM, J. G., CHEN, M.-H., HU, K. and JIA, C. (2014). Multivariate recurrent events in the presence of multivariate informative censoring with applications to bleeding and transfusion events in myelodysplastic syndrome. *J. Biopharm. Statist.* **24** 429–442. MR3196150 https://doi.org/10.1080/10543406.2013.860159

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 46–67
https://doi.org/10.1214/25-AOAS2116

INTEGRATIVE LEARNING OF LINEAR NON-GAUSSIAN DIRECTED ACYCLIC GRAPHS WITH APPLICATION ON MULTISOURCE GENE REGULATORY NETWORK ANALYSIS

BY XUANYU LI[1,2,a], SANGUO ZHANG[1,2,b], MINGYANG REN[3,c] AND QINGZHAO ZHANG[4,d]

[1]*School of Mathematical Sciences, University of Chinese Academy of Sciences*

[2]*Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences,* [a]*lixuanyu22@mails.ucas.ac.cn,* [b]*sgzhang@ucas.ac.cn*

[3]*School of Mathematical Sciences, Shanghai Jiao Tong University,* [c]*mingyangren@sjtu.edu.cn*

[4]*School of Economics and The Wang Yanan Institute for Studies in Economics, Xiamen University,* [d]*qzzhang@xmu.edu.cn*

A Directed Acyclic Graph (DAG) is a fundamental model for representing directional relationships among a set of random variables, with extensive applications in biology and medicine. Yet the limited data in one single study may affect accurate DAG reconstruction, whereas data from multiple relevant studies can be collected. It raises the challenging question of how to integrate multiple studies for better constructing common DAG structures. In this article we consider multiple linear non-Gaussian DAGs in high-dimensional cases and propose a novel integrative learning framework. Our framework requires only that multiple DAGs share a common structure but can have specific edge strengths and noise distributions. We also establish the asymptotic consistency result in terms of the DAG reconstruction, which shows substantial theoretical improvement of the integrative DAG learning in multiple aspects compared to the single DAG learning. The advantage of our proposed method is further supported by the numerical comparison of synthetic data as well as multisite nonsmall cell lung cancer data.

REFERENCES

ASSOUN, S., THEOU-ANTON, N., NGUENANG, M., CAZES, A., DANEL, C., ABBAR, B., PLUVY, J., GOUNANT, V., KHALIL, A. et al. (2019). Association of TP53 mutations with response and longer survival under immune checkpoint inhibitors in advanced non-small-cell lung cancer. *Lung Cancer* **132** 65–71. https://doi.org/10.1016/j.lungcan.2019.04.005

BILGRAU, A. E., PEETERS, C. F. W., ERIKSEN, P. S., BØGSTED, M. and VAN WIERINGEN, W. N. (2020). Targeted fused ridge estimation of inverse covariance matrices from multiple high-dimensional data classes. *J. Mach. Learn. Res.* **21** Paper No. 26, 52. MR4073759

BOLSTAD, B. M., IRIZARRY, R. A., ÅSTRAND, M. and SPEED, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19** 185–193.

CAI, T., LIU, M. and XIA, Y. (2022). Individual data protected integrative regression analysis of high-dimensional heterogeneous data. *J. Amer. Statist. Assoc.* **117** 2105–2119. MR4528492 https://doi.org/10.1080/01621459.2021.1904958

CANALE, M., PETRACCI, E., DELMONTE, A., CHIADINI, E., DAZZI, C., PAPI, M., CAPELLI, L., CASANOVA, C., DE LUIGI, N. et al. (2017). Impact of TP53 mutations on outcome in EGFR-mutated patients treated with first-line tyrosine kinase inhibitors. *Clin. Cancer Res.* **23** 2195–2202. https://doi.org/10.1158/1078-0432.CCR-16-0966

CASTELLETTI, F., CONSONNI, G., DELLA VEDOVA, M. L. and PELUSO, S. (2018). Learning Markov equivalence classes of directed acyclic graphs: An objective Bayes approach. *Bayesian Anal.* **13** 1235–1260. MR3855370 https://doi.org/10.1214/18-BA1101

CHEN, T., BELLO, K., ARAGAM, B. and RAVIKUMAR, P. (2023). iSCAN: Identifying causal mechanism shifts among nonlinear additive noise models. *Adv. Neural Inf. Process. Syst.* **36** 44671–44706.

Key words and phrases. Directed acyclic graph, integrative analysis, permutation test, structural equation model, topological layer.

CHEN, T., BELLO, K., LOCATELLO, F., ARAGAM, B. and RAVIKUMAR, P. (2024). Identifying general mechanism shifts in linear causal representations. *Adv. Neural Inf. Process. Syst.* **37** 42405–42429.

CHEN, X., SUN, H., ELLINGTON, C., XING, E. and SONG, L. (2021). Multi-task learning of order-consistent causal graphs. *Adv. Neural Inf. Process. Syst.* **34** 11083–11095.

CHICKERING, D. M. (2002). Optimal structure identification with greedy search. *J. Mach. Learn. Res.* **3** 507–554. MR1991085 https://doi.org/10.1162/153244303321897717

CHOWDHURY, S., WANG, R., YU, Q., HUNTOON, C. J., KARNITZ, L. M., KAUFMANN, S. H., GYGI, S. P., BIRRER, M. J., PAULOVICH, A. G. et al. (2022). DAGBagM: Learning directed acyclic graphs of mixed variables with an application to identify protein biomarkers for treatment response in ovarian cancer. *BMC Bioinform.* **23** 321. https://doi.org/10.1186/s12859-022-04864-y

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

DE TORRENTÉ, L., ZIMMERMAN, S., SUZUKI, M., CHRISTOPEIT, M., GREALLY, J. M. and MAR, J. C. (2020). The shape of gene expression distributions matter: How incorporating distribution shape improves the interpretation of cancer transcriptomic data. *BMC Bioinform.* **21** 1–18.

GAO, M., DING, Y. and ARAGAM, B. (2020). A polynomial-time algorithm for learning nonparametric causal graphs. *Adv. Neural Inf. Process. Syst.* **33** 11599–11611.

GHOSHAL, A., BELLO, K. and HONORIO, J. (2019). Direct learning with guarantees of the difference dag between structural equation models. arXiv preprint. Available at arXiv:1906.12024.

GHOSHAL, A. and HONORIO, J. (2018). Learning linear structural equation models in polynomial time and sample complexity. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics* (A. Storkey and F. Perez-Cruz, eds.). *Proceedings of Machine Learning Research* **84** PMLR.

HA, M. J., SUN, W. and XIE, J. (2016). `PenPC`: A two-step approach to estimate the skeletons of high-dimensional directed acyclic graphs. *Biometrics* **72** 146–155. MR3500583 https://doi.org/10.1111/biom.12415

HUANG, Y., ZHANG, Q., ZHANG, S., HUANG, J. and MA, S. (2017). Promoting similarity of sparsity structures in integrative analysis with penalization. *J. Amer. Statist. Assoc.* **112** 342–350. MR3646576 https://doi.org/10.1080/01621459.2016.1139497

HUO, X. and SZÉKELY, G. J. (2016). Fast computing for distance covariance. *Technometrics* **58** 435–447. MR3556612 https://doi.org/10.1080/00401706.2015.1054435

KALISCH, M. and BÜHLMANN, P. (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *J. Mach. Learn. Res.* **8** 613–636.

LENG, C., LIN, Y. and WAHBA, G. (2006). A note on the lasso and related procedures in model selection. *Statist. Sinica* **16** 1273–1284. MR2327490

LI, X., ZHANG, S., REN, M. and ZHANG, Q. (2026). Supplement to "Integrative learning of linear non-Gaussian directed acyclic graphs with application on multisource gene regulatory network analysis." https://doi.org/10.1214/25-AOAS2116SUPPA, https://doi.org/10.1214/25-AOAS2116SUPPB

LI, Y., GUESSOUS, F., KWON, S., KUMAR, M., IBIDAPO, O., FULLER, L., JOHNSON, E., LAL, B., HUSSAINI, I. et al. (2008). PTEN has tumor-promoting properties in the setting of gain-of-function p53 mutations. *Cancer Res.* **68** 1723–1731.

LIU, J., SUN, W. and LIU, Y. (2019). Joint skeleton estimation of multiple directed acyclic graphs for heterogeneous population. *Biometrics* **75** 36–47. MR3953705 https://doi.org/10.1111/biom.12941

LIU, L., WANG, G., WANG, L., YU, C., LI, M., SONG, S., HAO, L., MA, L. and ZHANG, Z. (2020). Computational identification and characterization of glioma candidate biomarkers through multi-omics integrative profiling. *Biol. Direct* **15** 10.

LIU, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. MR3161453 https://doi.org/10.1214/13-AOS1169

LIU, X., CHEN, L., TIAN, X. and ZHANG, T. (2017). MiR-137 and its target TGFA modulate cell growth and tumorigenesis of non-small cell lung cancer. *Eur. Rev. Med. Pharmacol. Sci.* **21** 511–517.

LOH, P.-L. and BÜHLMANN, P. (2014). High-dimensional learning of linear causal networks via inverse covariance estimation. *J. Mach. Learn. Res.* **15** 3065–3105. MR3277162

LOUNICI, K., PONTIL, M., TSYBAKOV, A. B. and VAN DE GEER, S. (2009). Taking advantage of sparsity in multi-task learning. arXiv preprint. Available at arXiv:0903.1468.

MADAN BABU, M. and TEICHMANN, S. A. (2003). Evolution of transcription factors and the gene regulatory network in Escherichia coli. *Nucleic Acids Res.* **31** 1234–1244.

MAJUMDAR, S. and MICHAILIDIS, G. (2022). Joint estimation and inference for data integration problems based on multiple multi-layered Gaussian graphical models. *J. Mach. Learn. Res.* **23** Paper No. 1, 53. MR4420726

MALIK, V., BELLO, K., GHOSHAL, A. and HONORIO, J. (2024). Identifying causal changes between linear structural equation models. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence* (N. Kiyavash and J. M. Mooij, eds.). *Proceedings of Machine Learning Research* **244** 2383–2398. PMLR.

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

MIZUARAI, S., MACHIDA, T., KOBAYASHI, T., KOMATANI, H., ITADANI, H. and KOTANI, H. (2011). Expression ratio of CCND1 to CDKN2A mRNA predicts RB1 status of cultured cancer cell lines and clinical tumor samples. *Mol. Cancer* **10** 31. https://doi.org/10.1186/1476-4598-10-31

OATES, C. J., SMITH, J. Q., MUKHERJEE, S. and CUSSENS, J. (2016). Exact estimation of multiple directed acyclic graphs. *Stat. Comput.* **26** 797–811. MR3515022 https://doi.org/10.1007/s11222-015-9570-9

PARK, G. and KIM, Y. (2020). Identifiability of Gaussian linear structural equation models with homogeneous and heterogeneous error variances. *J. Korean Statist. Soc.* **49** 276–292. MR4122465 https://doi.org/10.1007/s42952-019-00019-7

PETERS, J. and BÜHLMANN, P. (2014). Identifiability of Gaussian structural equation models with equal error variances. *Biometrika* **101** 219–228. MR3180667 https://doi.org/10.1093/biomet/ast043

REN, M., HE, X. and WANG, J. (2023). Structural transfer learning of non-Gaussian DAG. arXiv preprint. Available at arXiv:2310.10239.

SHEDDEN, K., TAYLOR, J. M. G., ENKEMANN, S. A., TSAO, M.-S., YEATMAN, T. J., GERALD, W. L., ESCHRICH, S., JURISICA, I., GIORDANO, T. J., MISEK, D. E., CHANG, A. C., ZHU, C. Q., STRUMPF, D., HANASH, S., SHEPHERD, F. A., DING, K., SEYMOUR, L., NAOKI, K., PENNELL, N., WEIR, B., VERHAAK, R., LADD-ACOSTA, C., GOLUB, T., GRUIDL, M., SHARMA, A., SZOKE, J., ZAKOWSKI, M., RUSCH, V., KRIS, M., VIALE, A., MOTOI, N., TRAVIS, W., CONLEY, B., SESHAN, V. E., MEYERSON, M., KUICK, R., DOBBIN, K. K., LIVELY, T., JACOBSON, J. W., BEER, D. G. and DIRECTOR'S CHALLENGE CONSORTIUM FOR THE MOLECULAR CLASSIFICATION OF LUNG ADENOCARCINOMA (2008). Gene expression–based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. *Nat. Med.* **14** 822–827. https://doi.org/10.1038/nm.1790

SHIMIZU, S. (2012). Joint estimation of linear non-Gaussian acyclic models. *Neurocomputing* **81** 104–107.

SHIMIZU, S., INAZUMI, T., SOGAWA, Y., HYVÄRINEN, A., KAWAHARA, Y., WASHIO, T., HOYER, P. O. and BOLLEN, K. (2011). DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *J. Mach. Learn. Res.* **12** 1225–1248. MR2804599

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2001). *Causation, Prediction, and Search*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. With additional material by David Heckerman, Christopher Meek, Gregory F. Cooper and Thomas Richardson. A Bradford Book. MR1815675

SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

TRAMONTANO, D., MONOD, A. and DRTON, M. (2022). Learning linear non-Gaussian polytree models. In *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence* (J. Cussens and K. Zhang, eds.). *Proceedings of Machine Learning Research* **180** 1960–1969. PMLR.

TSAMARDINOS, I., BROWN, L. E. and ALIFERIS, C. F. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Mach. Learn.* **65** 31–78.

VOGELSTEIN, B., PAPADOPOULOS, N., VELCULESCU, V. E., ZHOU, S., DIAZ, L. A. and KINZLER, K. W. (2013). Cancer genome landscapes. *Science* **339** 1546–1558. https://doi.org/10.1126/science.1235122

WANG, Y., SEGARRA, S. and UHLER, C. (2020). High-dimensional joint estimation of multiple directed Gaussian graphical models. *Electron. J. Stat.* **14** 2439–2483. MR4118334 https://doi.org/10.1214/20-EJS1724

WANG, Y., SQUIRES, C., BELYAEVA, A. and UHLER, C. (2018). Direct estimation of differences in causal graphs. In *Advances in Neural Information Processing Systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds.). **31**. Curran Associates.

WANG, Y. S. and DRTON, M. (2020). High-dimensional causal discovery under non-Gaussianity. *Biometrika* **107** 41–59. MR4064139 https://doi.org/10.1093/biomet/asz055

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

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

ZHAO, R., HE, X. and WANG, J. (2022). Learning linear non-Gaussian directed acyclic graph with diverging number of nodes. *J. Mach. Learn. Res.* **23** Paper No. [269], 34. MR4577708

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 68–89
https://doi.org/10.1214/26-AOAS2136

BRIDGING THE GAP: ENHANCING THE GENERALIZABILITY OF EPIGENETIC CLOCKS THROUGH TRANSFER LEARNING

BY LAN LUO[1,a], LULU SHANG[2,b], JACLYN M. GOODRICH[3,c], KAREN E. PETERSON[4,d] AND PETER X. K. SONG[5,e]

[1]*Department Biostatistics and Epidemiology, Rutgers University,* [a]*l.luo@rutgers.edu*
[2]*Department of Biostatistics, University of Texas MD Anderson Cancer Center,* [b]*LShang@mdanderson.org*
[3]*Department of Environmental Health Sciences, University of Michigan,* [c]*gaydojac@umich.edu*
[4]*Department of Nutritional Sciences, University of Michigan,* [d]*karenep@umich.edu*
[5]*Department of Biostatistics, University of Michigan,* [e]*pxsong@umich.edu*

Changes in DNA methylation patterns exhibit a high correlation with chronological age. Epigenetic clocks, developed through statistical models that estimate epigenetic age using the methylation levels of cytosine-guanine dinucleotide (CpG) sites, have emerged as powerful tools for understanding aging and age-related diseases. Despite their popularity, the generalizability of these clocks across diverse populations remains a challenge. Some of the widely used epigenetic clocks, such as Horvath's clock (*Genome Biol.* **14** (2013) 1–20) and the PedBE clock (*Proc. Natl. Acad. Sci. USA* **117** (2020) 23329–23335), are shown to perform poorly in our target cohort. This loss of prediction accuracy raises concerns about their viability in calculating biological age in distinct demographic and ethnic groups. Technically, the feature space of existing clocks is yielded with an obsolete technique, potentially leading to systematic bias in the analysis of all target data generated by the EPIC 850K array. To address both population heterogeneity and technological advances, we adopt a transfer learning framework to calibrate existing epigenetic clocks by borrowing shared knowledge from diverse datasets. Furthermore, our transfer learning is built on kriging- and DNN-based methods for feature adaptation, to close the gap between existing clocks and our target data. We analyze data collected from 523 blood samples from a cohort of children and adolescents in the Early Life Exposure in Mexico to Environmental Toxicants (ELEMENT) study and show that our proposed transfer learning methods significantly improve prediction performance compared to existing clocks. Performance is further enhanced by using the CpG sites profiled on the higher-resolution EPIC array. More importantly, calibrated clocks produce epigenetic age accelerations that correlate better with stages of sexual maturation. Our methodology demonstrates the potential to bridge the gap between different DNA methylation datasets and various profiling platforms, thereby enhancing the applicability of epigenetic clocks across diverse population groups and contributing to more accurate aging research.

REFERENCES

AANES, H., BLEKA, Ø., DAHLBERG, P. S., CARM, K. T., LEHTIMÄKI, T., RAITAKARI, O., KÄHÖNEN, M., HURME, M. and ROLSETH, V. (2023). A new blood based epigenetic age predictor for adolescents and young adults. *Sci. Rep.* **13** 2303.

ALMSTRUP, K., LINDHARDT JOHANSEN, M., BUSCH, A. S., HAGEN, C. P., NIELSEN, J. E., PETERSEN, J. H. and JUUL, A. (2016). Pubertal development in healthy children is mirrored by DNA methylation patterns in peripheral blood. *Sci. Rep.* **6** 28657.

BELSKY, D. W., CASPI, A., ARSENEAULT, L., BACCARELLI, A., CORCORAN, D. L., GAO, X., HANNON, E., HARRINGTON, H. L., RASMUSSEN, L. J. et al. (2020). Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. *eLife* **9** e54870.

Key words and phrases. Transfer learning, epigenetic clocks, epigenetic age acceleration, population heterogeneity, platform heterogeneity.

Belsky, D. W., Caspi, A., Corcoran, D. L., Sugden, K., Poulton, R., Arseneault, L., Baccarelli, A., Chamarti, K., Gao, X. et al. (2022). DunedinPACE, a DNA methylation biomarker of the pace of aging. *eLife* **11** e73420.

Cathey, A., Watkins, D. J., Sánchez, B. N., Tamayo-Ortiz, M., Solano-Gonzalez, M., Torres-Olascoaga, L., Téllez-Rojo, M. M., Peterson, K. E. and Meeker, J. D. (2020). Onset and tempo of sexual maturation is differentially associated with gestational phthalate exposure between boys and girls in a Mexico City birth cohort. *Environ. Int.* **136** 105469.

Chen, B. H., Marioni, R. E., Colicino, E., Peters, M. J., Ward-Caviness, C. K., Tsai, P.-C., Roetker, N. S., Just, A. C., Demerath, E. W. et al. (2016). DNA methylation-based measures of biological age: Meta-analysis predicting time to death. *Aging* **8** 1844.

Chinn, I. K., Blackburn, C. C., Manley, N. R. and Sempowski, G. D. (2012). Changes in primary lymphoid organs with aging. In *Seminars in Immunology* **24** 309–320. Elsevier, Amsterdam.

Cortez, B. N., Pan, H., Hinthorn, S., Sun, H., Neretti, N., Gloyn, A. L. and Aguayo-Mazzucato, C. (2024). Heterogeneity of increased biological age in type 2 diabetes correlates with differential tissue DNA methylation, biological variables, and pharmacological treatments. *GeroScience* **46** 2441–2461.

Cressie, N. A. C. (2015). *Statistics for Spatial Data*, Revised ed. *Wiley Classics Library*. Wiley, New York. Paperback edition of the 1993 edition [MR1239641]. MR3559472

de Lima Camillo, L. P., Lapierre, L. R. and Singh, R. (2022). A pan-tissue DNA-methylation epigenetic clock based on deep learning. *npj Aging* **8**.

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R. and Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10** 3328.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V. K., Attwood, J., Burger, M., Burton, J., Cox, T. V., Davies, R. et al. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* **38** 1378–1385.

Ehlinger, J. V., Goodrich, J. M., Dolinoy, D. C., Watkins, D. J., Cantoral, A., Mercado-García, A., Téllez-Rojo, M. M. and Peterson, K. E. (2023). Associations between blood leukocyte DNA methylation and sustained attention in mid-to-late childhood. *Epigenomics* **15** 965–981.

Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x

Faul, J. D., Kim, J. K., Levine, M. E., Thyagarajan, B., Weir, D. R. and Crimmins, E. M. (2023). Epigenetic-based age acceleration in a representative sample of older Americans: Associations with aging-related morbidity and mortality. *Proc. Natl. Acad. Sci. USA* **120**. e2215840120.

FC Lopes, A. (2020). Mitochondrial metabolism and DNA methylation: A review of the interaction between two genomes. *Clin. Epigenet.* **12** 182.

Fortin, J.-P., Triche, T. J. Jr and Hansen, K. D. (2017). Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33** 558–560.

Galkin, F., Mamoshina, P., Kochetov, K., Sidorenko, D. and Zhavoronkov, A. (2021). DeepMAge: A methylation aging clock developed with deep learning. *Aging Dis.* **12** 1252.

Gu, T., Han, Y. and Duan, R. (2025). Robust angle-based transfer learning in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **87** 723–745. MR4941743 https://doi.org/10.1093/jrsssb/qkae111

Halabicky, O., Téllez-Rojo, M., Goodrich, J., Dolinoy, D., Mercado-García, A., Hu, H. and Peterson, K. (2024). Prenatal and childhood lead exposure is prospectively associated with biological markers of aging in adolescence. *Sci. Total Environ.* **913** 169757.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.-B. et al. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49** 359–367.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biol.* **14** 1–20.

Horvath, S., Oshima, J., Martin, G. M., Lu, A. T., Quach, A., Cohen, H., Felton, S., Matsuyama, M., Lowe, D. et al. (2018). Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging* **10** 1758.

Jain, P., Binder, A. M., Chen, B., Parada, H., Gallo, L. C., Alcaraz, J., Horvath, S., Bhatti, P., Whitsel, E. A. et al. (2022). Analysis of epigenetic age acceleration and healthy longevity among older US women. *JAMA Netw. Open* **5** e2223285–e2223285.

Jansen, E. C., Dolinoy, D., Peterson, K. E., O'Brien, L. M., Chervin, R. D., Cantoral, A., Tellez-Rojo, M. M., Solano-Gonzalez, M. and Goodrich, J. (2021). Adolescent sleep timing and dietary patterns in relation to DNA methylation of core circadian genes: A pilot study of Mexican youth. *Epigenetics* **16** 894–907.

Joseph, V. R. (2022). Optimal ratio for data splitting. *Stat. Anal. Data Min.* **15** 531–538. MR4461844 https://doi.org/10.1002/sam.11583

KIM, C., HARRALL, K. K., GLUECK, D. H., HOCKETT, C. and DABELEA, D. (2024). Epigenetic age acceleration is associated with speed of pubertal growth but not age of pubertal onset. *Sci. Rep.* **14** 2981.

KRAFT, S. A., CHO, M. K., GILLESPIE, K., HALLEY, M., VARSAVA, N., ORMOND, K. E., LUFT, H. S., WILFOND, B. S. and LEE, S. S.-J. (2018). Beyond consent: Building trusting relationships with diverse populations in precision medicine research. *Amer. J. Bioeth.* **18** 3–20.

LANDRY, L. G., ALI, N., WILLIAMS, D. R., REHM, H. L. and BONHAM, V. L. (2018). Lack of diversity in genomic databases is a barrier to translating precision medicine research into practice. *Health Aff.* **37** 780–785.

LEVINE, M. E., LU, A. T., QUACH, A., CHEN, B. H., ASSIMES, T. L., BANDINELLI, S., HOU, L., BACCARELLI, A. A., STEWART, J. D. et al. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10** 573.

LEVY, J. J., TITUS, A. J., PETERSEN, C. L., CHEN, Y., SALAS, L. A. and CHRISTENSEN, B. C. (2020). MethylNet: An automated and modular deep learning approach for DNA methylation analysis. *BMC Bioinform.* **21** 1–15.

LI, S., CAI, T. T. and LI, H. (2022). Transfer learning for high-dimensional linear regression: Prediction, estimation and minimax optimality. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 149–173. MR4400393

LI, Z., SHEN, Y. and NING, J. (2023). Accommodating time-varying heterogeneity in risk estimation under the Cox model: A transfer learning approach. *J. Amer. Statist. Assoc.* **118** 2276–2287. MR4681582 https://doi.org/10.1080/01621459.2023.2210336

LU, A. T., FEI, Z., HAGHANI, A., ROBECK, T. R., ZOLLER, J., LI, C., LOWE, R., YAN, Q., ZHANG, J. et al. (2023). Universal DNA methylation age across mammalian tissues. *Nat. Aging* **3** 1144–1166.

LU, A. T., QUACH, A., WILSON, J. G., REINER, A. P., AVIV, A., RAJ, K., HOU, L., BACCARELLI, A. A., LI, Y. et al. (2019). DNA methylation-based estimator of telomere length. *Aging* **11** 5895.

MARIONI, R. E., SHAH, S., MCRAE, A. F., CHEN, B. H., COLICINO, E., HARRIS, S. E., GIBSON, J., HENDERS, A. K., REDMOND, P. et al. (2015). DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* **16** 1–12.

MATHERON, G. (1963). Principles of geostatistics. *Econ. Geol.* **58** 1246–1266.

MCEWEN, L. M., O'DONNELL, K. J., MCGILL, M. G., EDGAR, R. D., JONES, M. J., MACISAAC, J. L., LIN, D. T. S., RAMADORI, K., MORIN, A. et al. (2020). The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. *Proc. Natl. Acad. Sci. USA* **117** 23329–23335.

MUNAFÒ, M. R., TILLING, K., TAYLOR, A. E., EVANS, D. M. and DAVEY SMITH, G. (2018). Collider scope: When selection bias can substantially influence observed associations. *Int. J. Epidemiol.* **47** 226–235.

NWANAJI-ENWEREM, J. C., VAN DER LAAN, L., KOGUT, K., ESKENAZI, B., HOLLAND, N., DEARDORFF, J. and CARDENAS, A. (2021). Maternal adverse childhood experiences before pregnancy are associated with epigenetic aging changes in their children. *Aging* **13** 25653.

PERNG, W., TAMAYO-ORTIZ, M., TANG, L., SÁNCHEZ, B. N., CANTORAL, A., MEEKER, J. D., DOLINOY, D. C., ROBERTS, E. F., MARTINEZ-MIER, E. A. et al. (2019). Early Life Exposure in Mexico to ENvironmental Toxicants (ELEMENT) project. *BMJ Open* **9** e030427.

PIDSLEY, R., ZOTENKO, E., PETERS, T. J., LAWRENCE, M. G., RISBRIDGER, G. P., MOLLOY, P., VAN DJIK, S., MUHLHAUSLER, B., STIRZAKER, C. et al. (2016). Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* **17** 1–17.

REYNOLDS, L. M., TAYLOR, J. R., DING, J., LOHMAN, K., JOHNSON, C., SISCOVICK, D., BURKE, G., POST, W., SHEA, S. et al. (2014). Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nat. Commun.* **5** 5366.

SHIREBY, G. L., DAVIES, J. P., FRANCIS, P. T., BURRAGE, J., WALKER, E. M., NEILSON, G. W., DAHIR, A., THOMAS, A. J., LOVE, S. et al. (2020). Recalibrating the epigenetic clock: Implications for assessing biological age in the human cortex. *Brain* **143** 3763–3775.

TAYLOR, A. E., JONES, H. J., SALLIS, H., EUESDEN, J., STERGIAKOULI, E., DAVIES, N. M., ZAMMIT, S., LAWLOR, D. A., MUNAFÒ, M. R. et al. (2018). Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **47** 1207–1216.

TESCHENDORFF, A. E. and HORVATH, S. (2025). Epigenetic ageing clocks: Statistical methods and emerging computational challenges. *Nat. Rev. Genet.* 1–19.

TIAN, Y. and FENG, Y. (2023). Transfer learning under high-dimensional generalized linear models. *J. Amer. Statist. Assoc.* **118** 2684–2697. MR4681613 https://doi.org/10.1080/01621459.2022.2071278

TIAN, Y. E., CROPLEY, V., MAIER, A. B., LAUTENSCHLAGER, N. T., BREAKSPEAR, M. and ZALESKY, A. (2023). Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. *Nat. Med.* **29** 1221–1231.

TYRRELL, J., ZHENG, J., BEAUMONT, R., HINTON, K., RICHARDSON, T. G., WOOD, A. R., DAVEY SMITH, G., FRAYLING, T. M. and TILLING, K. (2021). Genetic predictors of participation in optional components of UK Biobank. *Nat. Commun.* **12** 886.

UNNIKRISHNAN, A., FREEMAN, W. M., JACKSON, J., WREN, J. D., PORTER, H. and RICHARDSON, A. (2019). The role of DNA methylation in epigenetics of aging. *Pharmacol. Ther.* **195** 172–185.

WATKINS, S. H., TESTA, C., CHEN, J. T., DE VIVO, I., SIMPKIN, A. J., TILLING, K., DIEZ ROUX, A. V., DAVEY SMITH, G., WATERMAN, P. D. et al. (2023). Epigenetic clocks and research implications of the lack of data on whom they have been developed: A review of reported and missing sociodemographic characteristics. *Environ*. *Epigenet*. **9**. dvad005.

WEST, K. M., BLACKSHER, E. and BURKE, W. (2017). Genomics, health disparities, and missed opportunities for the nation's research agenda. *JAMA* **317** 1831–1832.

WU, X., CHEN, W., LIN, F., HUANG, Q., ZHONG, J., GAO, H., SONG, Y. and LIANG, H. (2019). DNA methylation profile is a quantitative measure of biological aging in children. *Aging* **11** 10031–10051. https://doi.org/10.18632/aging.102399

XU, Z., LANGIE, S. A., DE BOEVER, P., TAYLOR, J. A. and NIU, L. (2017). RELIC: A novel dye-bias correction method for Illumina Methylation BeadChip. *BMC Genomics* **18** 1–7.

ZHANG, Z. (2018). Improved Adam optimizer for deep neural networks. In 2018 *IEEE/ACM* 26*th International Symposium on Quality of Service* (*IWQoS*) 1–2. https://doi.org/10.1109/IWQoS.2018.8624183

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **67** 301–320. MR2137327 https://doi.org/10.1111/j.1467-9868.2005.00503.x

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 90–109
https://doi.org/10.1214/26-AOAS2145

MODEL-FREE INFERENCE FOR CHARACTERIZING PROTEIN MUTATIONS THROUGH A COEVOLUTIONARY LENS

BY FAN F. YANG[1,a], ZHAO REN[1,b], WEN ZHOU[2,c], KEJUE JIA[3,d] AND ROBERT JERNIGAN[4,e]

[1]*Department of Statistics, University of Pittsburgh,* [a]*ffy1@pitt.edu,* [b]*zren@pitt.edu*
[2]*Department of Biostatistics, School of Global Public Health, New York University,* [c]*wz3030@nyu.edu*
[3]*Department of Molecular, Cellular and Developmental Biology, Yale University,* [d]*kejue.jia@yale.edu*
[4]*Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University,* [e]*jernigan@iastate.edu*

Multiple sequence alignment (MSA) data play a crucial role in the study of protein mutations, with contact prediction being a notable application. Existing methods are often model-based or algorithmic and typically do not incorporate statistical inference to quantify the uncertainty of the prediction outcomes. To address this, we propose a novel framework that transforms the task of contact prediction into a statistical testing problem. Our approach is motivated by the partial correlation for continuous random variables. With one-hot encoding of MSA data, we are able to construct a partial correlation graph for multivariate categorical variables. In this framework, two connected nodes in the graph indicate that the corresponding positions on the protein form a contact. A new spectrum-based test statistic is introduced to test whether two positions are partially correlated. Moreover, the new framework enables the identification of amino acid combinations that contribute to the correlation within the identified contacts, an important but largely unexplored aspect of protein mutations. Numerical experiments demonstrate that our proposed method is valid in terms of controlling Type I errors and powerful in general. Real data applications on various protein families further validate the practical utility of our approach in coevolution and mutation analysis.

REFERENCES

ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. *Wiley Publications in Statistics*. Wiley, New York. MR0091588

ASHKENAZY, H. and KLIGER, Y. (2010). Reducing phylogenetic bias in correlated mutation analysis. *Protein Eng. Des. Sel.* **23** 321–326.

BALDASSI, C., ZAMPARO, M., FEINAUER, C., PROCACCINI, A., ZECCHINA, R., WEIGT, M. and PAGNANI, A. (2014). Fast and accurate multivariate Gaussian modeling of protein families: Predicting residue contacts and protein-interaction partners. *PLoS ONE* **9** e92721.

CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618 https://doi.org/10.1080/01621459.2012.758041

CAI, T. T. and LIU, W. (2016). Large-scale multiple testing of correlations. *J. Amer. Statist. Assoc.* **111** 229–240. MR3494655 https://doi.org/10.1080/01621459.2014.999157

CAPORASO, J. G., SMIT, S., EASTON, B. C., HUNTER, L., HUTTLEY, G. A. and KNIGHT, R. (2008). Detecting coevolution without phylogenetic trees? Tree-ignorant metrics of coevolution perform as well as tree-aware metrics. *BMC Evol. Biol.* **8** 327.

CHANG, J., QIU, Y., YAO, Q. and ZOU, T. (2018). Confidence regions for entries of a large precision matrix. *J. Econometrics* **206** 57–82. MR3840783 https://doi.org/10.1016/j.jeconom.2018.03.020

DENG, X. and YUAN, M. (2009). Large Gaussian covariance matrix estimation with Markov structures. *J. Comput. Graph. Statist.* **18** 640–657. MR2751644 https://doi.org/10.1198/jcgs.2009.07170

DUNN, S. D., WAHL, L. M. and GLOOR, G. B. (2008). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics* **24** 333–340.

Key words and phrases. Multivariate categorical data, multiple sequence alignment, partial correlation, precision matrix, protein mutation.

Ekeberg, M., Lövkvist, C., Lan, Y., Weigt, M. and Aurell, E. (2013). Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. *Phys. Rev. E, Stat. Nonlin. Soft Matter Phys.* **87** 012707.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Matloob, Q., Lorna, J. R., Gustavo, A. S. et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* **47** D427–D432.

Figliuzzi, M., Jacquier, H., Schug, A., Tenaillon, O. and Weigt, M. (2016). Coevolutionary landscape inference and the context-dependence of mutations in beta-lactamase TEM-1. *Mol. Biol. Evol.* **33** 268–280.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R. et al. (2014). Pfam: The protein families database. *Nucleic Acids Res.* **42** D222–D230.

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

Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

Hopf, T. A., Ingraham, J. B., Poelwijk, F. J., Schärfe, C. P., Springer, M., Sander, C. and Marks, D. S. (2017). Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **35** 128–135.

Izenman, A. J. (2008). *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. Springer Texts in Statistics.* Springer, New York. MR2445017 https://doi.org/10.1007/978-0-387-78189-1

Janková, J. and van de Geer, S. (2017). Honest confidence regions and optimality in high-dimensional precision matrix estimation. *TEST* **26** 143–162. MR3613609 https://doi.org/10.1007/s11749-016-0503-5

Jernigan, R., Jia, K., Ren, Z. and Zhou, W. (2021). Large-scale multiple inference of collective dependence with applications to protein function. *Ann. Appl. Stat.* **15** 902–924. MR4298967 https://doi.org/10.1214/20-aoas1431

Jia, K. and Jernigan, R. L. (2021). New amino acid substitution matrix brings sequence alignments into agreement with structure matches. *Proteins, Struct. Funct. Bioinform.* **89** 671–682.

Jia, K., Kilinc, M. and Jernigan, R. L. (2023). Functional protein dynamics directly from sequences. *J. Phys. Chem., B* **127** 1914–1921.

Jones, D. T., Buchan, D. W., Cozzetto, D. and Pontil, M. (2012). PSICOV: Precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics* **28** 184–190.

Kamisetty, H., Ovchinnikov, S. and Baker, D. (2013). Assessing the utility of coevolution-based residue–residue contact predictions in a sequence-and structure-rich era. *Proc. Natl. Acad. Sci. USA* **110** 15674–15679.

Lee, K.-Y., Ji, D., Li, L., Constable, T. and Zhao, H. (2023). Conditional functional graphical models. *J. Amer. Statist. Assoc.* **118** 257–271. MR4571120 https://doi.org/10.1080/01621459.2021.1924178

Li, Y., Nan, B. and Zhu, J. (2015). Multivariate sparse group lasso for the multivariate multiple linear regression with an arbitrary group structure. *Biometrics* **71** 354–363. MR3366240 https://doi.org/10.1111/biom.12292

Lin, Z., Akin, H., Rao, R. et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379** 1123–1130. MR4567681

Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. MR3161453 https://doi.org/10.1214/13-AOS1169

Liu, Y. and Bahar, I. (2012). Sequence evolution correlates with structural dynamics. *Mol. Biol. Evol.* **29** 2253–2263.

Marks, D. S., Hopf, T. A. and Sander, C. (2012). Protein structure prediction from sequence variation. *Nat. Biotechnol.* **30** 1072–1080.

Meier, J., Rao, R., Verkuil, R., Liu, J., Sercu, T. and Rives, A. (2021). Language models enable zero-shot prediction of the effects of mutations on protein function. *Adv. Neural Inf. Process. Syst.* **34** 29287–29303.

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

Mitra, R. and Zhang, C.-H. (2016). The benefit of group sparsity in group inference with de-biased scaled group Lasso. *Electron. J. Stat.* **10** 1829–1873. MR3522662 https://doi.org/10.1214/16-EJS1120

Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C. et al. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. USA* **108** E1293–E1301.

Muirhead, R. J. and Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika* **67** 31–43. MR0570502 https://doi.org/10.1093/biomet/67.1.31

Muller, K. E. and Peterson, B. L. (1984). Practical methods for computing power in testing the multivariate general linear hypothesis. *Comput. Statist. Data Anal.* **2** 143–158.

Murphy, L. R., Wallqvist, A. and Levy, R. M. (2000). Simplified amino acid alphabets for protein fold recognition and implications for folding. *Protein Eng.* **13** 149–152.

REN, M., ZHANG, S., ZHANG, Q. and MA, S. (2022). Gaussian graphical model-based heterogeneity analysis via penalized fusion. *Biometrics* **78** 524–535. MR4450573 https://doi.org/10.1111/biom.13426

REN, Z., SUN, T., ZHANG, C.-H. and ZHOU, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.* **43** 991–1026. MR3346695 https://doi.org/10.1214/14-AOS1286

RIVES, A., MEIER, J., SERCU, T., GOYAL, S., LIN, Z., LIU, J. et al. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. USA* **118**. e2016239118.

SCHMIRLER, R., HEINZINGER, M. and ROST, B. (2024). Fine-tuning protein language models boosts predictions across diverse tasks. *Nat. Commun.* **15** 7407.

STIFFLER, M. A., HEKSTRA, D. R. and RANGANATHAN, R. (2015). Evolvability as a function of purifying selection in TEM-1 β-lactamase. *Cell* **160** 882–892.

STORZ, J. F. (2018). Compensatory mutations and epistasis for protein function. *Curr. Opin. Struck. Biol.* **50** 18–25.

WEIGT, M., WHITE, R. A., SZURMANT, H., HOCH, J. A. and HWA, T. (2009). Identification of direct residue contacts in protein–protein interaction by message passing. *Proc. Natl. Acad. Sci. USA* **106** 67–72.

XIA, Y., CAI, T. and CAI, T. T. (2018). Multiple testing of submatrices of a precision matrix with applications to identification of between pathway interactions. *J. Amer. Statist. Assoc.* **113** 328–339. MR3803468 https://doi.org/10.1080/01621459.2016.1251930

YANG, F. F., REN, Z., ZHOU, W., JIA, K. and JERNIGAN, R. (2026). Supplement to “Model-Free Inference for Characterizing Protein Mutations through a Coevolutionary Lens.” https://doi.org/10.1214/26-AOAS2145SUPPA, https://doi.org/10.1214/26-AOAS2145SUPPB

ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

ZHANG, J. and LI, Y. (2023). High-dimensional Gaussian graphical regression models with covariates. *J. Amer. Statist. Assoc.* **118** 2088–2100. MR4646628 https://doi.org/10.1080/01621459.2022.2034632

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 110–130
https://doi.org/10.1214/25-AOAS2105
© Institute of Mathematical Statistics, 2026

ANALYSING DYNAMIC CROSS-PRICE DEPENDENCIES WITH A MARKOV-SWITCHING SPATIAL AUTOREGRESSIVE MODEL

BY MATTEO IACOPINI[1,a], TAMÁS KRISZTIN[2,b] AND PHILIPP PIRIBAUER[3,c]

[1]*Department of AI, Data and Decision Sciences, LUISS University,* [a]*miacopini@luiss.it*
[2]*Integrated Biosphere Futures (IBF) Research Group, International Institute for Applied Systems Analysis,* [b]*krisztin@iiasa.ac.at*
[3]*Regional Economics and Spatial Analysis, Austrian Institute of Economic Research,* [c]*philipp.piribauer@wifo.ac.at*

This study introduces a novel Markov-switching spatial autoregressive (MS-SAR) model to analyse dynamic cross-price interdependencies within the three-digit subcomponents of the Consumer Price Index (CPI) for 15 European Union countries. By allowing the spatial weight matrix and network strength to evolve over time, our model captures the complex, time-varying nature of economic interdependencies that traditional models often overlook. Our results reveal marked cross-country differences in the propagation of price shocks across different categories, providing valuable insights into the transmission of macroeconomic shocks, such as the recent energy price shock, to inflation dynamics.

REFERENCES

ALLEN, T. and ARKOLAKIS, C. (2023). Economic activity across space. *J. Econ. Perspect.* **37** 3–28.

ANSELIN, L. (1988). *Spatial Econometrics: Methods and Models* 4. Springer, Berlin.

BALTAGI, B. H., EGGER, P. H. and KESINA, M. (2016). Firm-level productivity spillovers in China's chemical industry: A spatial Hausman–Taylor approach. *J. Appl. Econometrics* **31** 214–248. MR3464857 https://doi.org/10.1002/jae.2460

BASILE, R., DURBÁN, M., MÍNGUEZ, R., MONTERO, J. M. and MUR, J. (2014). Modeling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities. *J. Econom. Dynam. Control* **48** 229–245. MR3274268 https://doi.org/10.1016/j.jedc.2014.06.011

BECK, G. W., HUBRICH, K. and MARCELLINO, M. (2016). On the importance of sectoral regional shocks for price-setting. *J. Appl. Econometrics* **31** 1234–1253. MR3580898 https://doi.org/10.1002/jae.2490

BILLIO, M., CASARIN, R., RAVAZZOLO, F. and VAN DIJK, H. K. (2016). Interconnections between Eurozone and US booms and busts using a Bayesian panel Markov-switching VAR model. *J. Appl. Econometrics* **31** 1352–1370. MR3580904 https://doi.org/10.1002/jae.2501

CARRIÈRE-SWALLOW, Y., DEB, P., FURCERI, D., JIMÉNEZ, D. and OSTRY, J. D. (2023). Shipping costs and inflation. *J. Int. Money Financ.* **130** 102771. https://doi.org/10.1016/j.jimonfin.2022.102771

CELEUX, G., FORBES, F., ROBERT, C. P. and TITTERINGTON, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Anal.* **1** 651–673. MR2282197 https://doi.org/10.1214/06-BA122

CORRADO, L. and FINGLETON, B. (2012). Where is the economics in spatial econometrics? *J. Reg. Sci.* **52** 210–239.

D'INNOCENZO, E., LUCAS, A., OPSCHOOR, A. and ZHANG, X. (2024). Heterogeneity and dynamics in network models. *J. Appl. Econometrics* **39** 150–173. MR4701961 https://doi.org/10.1002/jae.3013

DE CARVALHO, M., LEONELLI, M. and ROSSI, A. (2020). Tracking change-points in multivariate extremes. arXiv preprint. Available at arXiv:2011.05067.

DEBARSY, N. and LESAGE, J. (2018). Flexible dependence modeling using convex combinations of different types of connectivity structures. *Reg. Sci. Urban Econ.* **69** 48–68.

DEBARSY, N. and LESAGE, J. P. (2022). Bayesian model averaging for spatial autoregressive models based on convex combinations of different types of connectivity matrices. *J. Bus. Econom. Statist.* **40** 547–558. MR4410881 https://doi.org/10.1080/07350015.2020.1840993

DUNN, P. K. and SMYTH, G. K. (1996). Randomized quantile residuals. *J. Comput. Graph. Statist.* **5** 236–244.

FAN, T., LÜ, L., SHI, D. and ZHOU, T. (2021). Characterizing cycle structure in complex networks. *Commun. Phys.* **4** 1–9.

Key words and phrases. Bayesian spatial autoregressive model, cross-price dependencies, energy price shock, inflation dynamics in Europe, Markov switching, random spatial weights.

FORTUNATO, S. (2010). Community detection in graphs. *Phys. Rep.* **486** 75–174. MR2580414 https://doi.org/10.1016/j.physrep.2009.11.002

FRÜHWIRTH-SCHNATTER, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *J. Amer. Statist. Assoc.* **96** 194–209. MR1952732 https://doi.org/10.1198/016214501750333063

FRÜHWIRTH-SCHNATTER, S. (2006). *Finite Mixture and Markov Switching Models. Springer Series in Statistics*. Springer, New York. MR2265601

GASPERONI, F., LUATI, A., PACI, L. and D'INNOCENZO, E. (2023). Score-driven modeling of spatio-temporal data. *J. Amer. Statist. Assoc.* **118** 1066–1077. MR4595477 https://doi.org/10.1080/01621459.2021.1970571

GLOCKER, C. and PIRIBAUER, P. (2023). Propagation of Price Shocks to CPI Inflation: the role of Cross-Demand Dependencies WIFO working paper No. 663, Austrian Institute of Economic Research.

GLOCKER, C. and PIRIBAUER, P. (2025). Consumer preferences and inflation diffusion. *Macroecon. Dyn.* **29** e80.

GORODNICHENKO, Y., SHEREMIROV, V. and TALAVERA, O. (2018). Price setting in online markets: Does IT click? *J. Eur. Econ. Assoc.* **16** 1764–1811.

GORODNICHENKO, Y. and TALAVERA, O. (2017). Price setting in online markets: Basic facts, international comparisons, and cross-border integration. *Amer. Econ. Rev.* **107** 249–282.

HAUZENBERGER, N. and PFARRHOFER, M. (2021). Bayesian state-space modeling for analyzing heterogeneous network effects of US monetary policy. *Scand. J. Econ.* **123** 1261–1291.

HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

HORVATH, S. (2011). *Weighted Network Analysis: Applications in Genomics and Systems Biology*. Springer, New York.

HUBER, F. and FISCHER, M. M. (2018). A Markov switching factor-augmented VAR model for analyzing US business cycles and monetary policy. *Oxf. Bull. Econ. Stat.* **80** 575–604.

IACOPINI, M., KRISZTIN, T. and PIRIBAUER, P. (2026). Supplement to "Analysing dynamic cross-price dependencies with a Markov-Switching spatial autoregressive model." https://doi.org/10.1214/25-AOAS2105SUPP

JEAN, W. H. and HELMS, B. P. (1983). Geometric mean approximations. *J. Financ. Quant. Anal.* **18** 287–293.

KRISZTIN, T. and PIRIBAUER, P. (2022). A Bayesian approach for the estimation of weight matrices in spatial autoregressive models. *Spatial Econ. Anal.* 1–20.

LAM, C. and SOUZA, P. C. L. (2020). Estimation and selection of spatial weight matrix in a spatial lag model. *J. Bus. Econom. Statist.* **38** 693–710. MR4115427 https://doi.org/10.1080/07350015.2019.1569526

LESAGE, J. and PACE, R. K. (2009). *Introduction to Spatial Econometrics*. CRC Press, Boca Raton.

MUMTAZ, H. and SURICO, P. (2012). Evolving international inflation dynamics: World and country-specific factors. *J. Eur. Econ. Assoc.* **10** 716–734.

NEELY, C. J. and RAPACH, D. E. (2011). International comovements in inflation rates and country characteristics. *J. Int. Money Financ.* **30** 1471–1490.

NEWMAN, M. E. J. (2010). *Networks: An Introduction*. Oxford Univ. Press, Oxford. MR2676073 https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

PIRIBAUER, P. and CRESPO CUARESMA, J. (2016). Bayesian variable selection in spatial autoregressive models. *Spatial Econ. Anal.* **11** 457–479.

PIRIBAUER, P., GLOCKER, C. and KRISZTIN, T. (2023). Beyond distance: The spatial relationships of European regional economic growth. *J. Econom. Dynam. Control* **155** 104735.

RITTER, C. and TANNER, M. A. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the griddy-Gibbs sampler. *J. Amer. Statist. Assoc.* **87** 861–868.

SCIDÁ, D. (2023). Structural VAR and financial networks: A minimum distance approach to spatial modeling. *J. Appl. Econometrics* **38** 49–68. MR4550080 https://doi.org/10.1002/jae.2935

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 583–639. MR1979380 https://doi.org/10.1111/1467-9868.00353

SUN, D., TSUTAKAWA, R. K. and SPECKMAN, P. L. (1999). Posterior distribution of hierarchical models using CAR(1) distributions. *Biometrika* **86** 341–350. MR1705418 https://doi.org/10.1093/biomet/86.2.341

ZHANG, X. and YU, J. (2018). Spatial weights matrix selection and model averaging for spatial autoregressive models. *J. Econometrics* **203** 1–18. MR3758324 https://doi.org/10.1016/j.jeconom.2017.05.021

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 131–147
https://doi.org/10.1214/25-AOAS2122
© Institute of Mathematical Statistics, 2026

QUANTILED CONDITIONAL VARIANCE, SKEWNESS, AND KURTOSIS BY CORNISH–FISHER EXPANSION

BY NINGNING ZHANG[1,a] AND KE ZHU[2,b]

[1]*TravelSky Technology Limited,* [a]*zhangningning@travelsky.com.cn*
[2]*Department of Statistics and Actuarial Science, The University of Hong Kong,* [b]*mazhuke@hku.hk*

The conditional variance, skewness, and kurtosis play a central role in time series analysis. To learn the three conditional moments (CMs), the News Impact Curve (NIC) has been widely used. Since these CMs are unobserved, their NICs are typically assumed to have certain parametric forms and then learned within a parametric model, which accounts for the dynamics of all three CMs. However, this inevitably brings two issues: the risk of model misspecification and the instability of model estimation, where the latter issue results from a necessary nonlinear constraint (on the conditional skewness and kurtosis) that requires a complex restriction on the admission region of model parameters. To avoid the above two issues, we propose a novel method to estimate the three CMs via the so-called quantiled CMs (QCMs). Under certain high-level condition, we show the consistency of the QCMs. In an application to three major exchange rates, we give a data-driven method to propose the nonparametric NICs for the three CMs, based on the QCMs. Our obtained nonparametric NICs indicate that the existing parametric NICs for conditional skewness and kurtosis may miscapture the impact of large shocks (in absolute value).

REFERENCES

ANDREWS, D. W. K. (1988). Laws of large numbers for dependent nonidentically distributed random variables. *Econometric Theory* **4** 458–467. MR0985156 https://doi.org/10.1017/S0266466600013396

BALI, T. G., MO, H. and TANG, Y. (2008). The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR. *J. Bank. Financ.* **32** 269–282.

BARONE-ADESSI, G., GIANNOPOULOS, K. and VOSPER, L. (1999). VaR without correlations for nonlinear portfolios. *J. Futures Mark.* **19** 583–602.

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327. MR0853051 https://doi.org/10.1016/0304-4076(86)90063-1

CORNISH, E. A. and FISHER, R. A. (1938). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Stat.* **5** 307–320.

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 MANGANELLI, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *J. Bus. Econom. Statist.* **22** 367–381. MR2091566 https://doi.org/10.1198/073500104000000370

ENGLE, R. F. and NG, V. K. (1993). Measuring and testing the impact of news on volatility. *J. Finance* **48** 1749–1778.

ESCANCIANO, J. C. (2006). Goodness-of-fit tests for linear and nonlinear time series models. *J. Amer. Statist. Assoc.* **101** 531–541. MR2256173 https://doi.org/10.1198/016214505000001050

ESCANCIANO, J. C. and VELASCO, C. (2006). Generalized spectral tests for the martingale difference hypothesis. *J. Econometrics* **134** 151–185. MR2328319 https://doi.org/10.1016/j.jeconom.2005.06.019

FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods. Springer Series in Statistics*. Springer, New York. MR1964455 https://doi.org/10.1007/b97702

GLOSTEN, L. R., JAGANNATHAN, R. and RUNKLE, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance* **48** 1779–1801.

GU, S., KELLY, B. and XIU, D. (2020). Empirical asset pricing via machine learning. *Rev. Financ. Stud.* **33** 2223–2273.

Key words and phrases. Conditional moments, Cornish–Fisher expansion, news impact curve, quantiled conditional moments.

HARVEY, C. R. and SIDDIQUE, A. (1999). Autoregressive conditional skewness. *J. Financ. Quant. Anal.* **34** 465–487.

HARVEY, C. R. and SIDDIQUE, A. (2000). Conditional skewness in asset pricing tests. *J. Finance* **55** 1263–1295.

JONDEAU, E. and ROCKINGER, M. (2003). Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. *J. Econom. Dynam. Control* **27** 1699–1737. MR1981727 https://doi.org/10.1016/S0165-1889(02)00079-9

JONDEAU, E., ZHANG, Q. and ZHU, X. (2019). Average skewness matters. *J. Financ. Econ.* **134** 29–47.

KOENKER, R. and XIAO, Z. (2006). Quantile autoregression. *J. Amer. Statist. Assoc.* **101** 980–990. MR2324109 https://doi.org/10.1198/016214506000000672

KOENKER, R. and ZHAO, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory* **12** 793–813. MR1421404 https://doi.org/10.1017/S0266466600007167

KUESTER, K., MITTNIK, S. and PAOLELLA, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *J. Financ. Econom.* **4** 53–89.

LEE, Y. S. and LIN, T. K. (1992). Algorithm AS 269: High order Cornish–Fisher expansion. *J. R. Stat. Soc., Ser. C* **41** 233–240.

LEÓN, Á., RUBIO, G. and SERNA, G. (2005). Autoregresive conditional volatility, skewness and kurtosis. *Q. Rev. Econ. Finance* **45** 599–618.

LIU, Y. and WU, Y. (2011). Simultaneous multiple non-crossing quantile regression estimation using kernel constraints. *J. Nonparametr. Stat.* **23** 415–437. MR2801302 https://doi.org/10.1080/10485252.2010.537336

LJUNG, G. M. and BOX, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika* **65** 297–303.

MCNEIL, A. J. and FREY, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *J. Empir. Finance* **7** 271–300.

ROBINSON, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56** 931–954. MR0951762 https://doi.org/10.2307/1912705

TONG, H. (1978). *On a Threshold Model* (C. H. Chen, ed.). *Pattern Recognition and Signal Processing*. Sijthoff & Noordhoff, Amsterdam.

TSAY, R. S. (2005). *Analysis of Financial Time Series*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, Hoboken, NJ. MR2162112 https://doi.org/10.1002/0471746193

WHITE, H., KIM, T.-H. and MANGANELLI, S. (2015). VAR for VaR: Measuring tail dependence using multivariate regression quantiles. *J. Econometrics* **187** 169–188. MR3347301 https://doi.org/10.1016/j.jeconom.2015.02.004

WIDDER, D. V. (1946). *The Laplace Transform*. *Princeton Mathematical Series*. Princeton Univ. Press, Princeton, NJ.

XIAO, Z. (2012). Time series quantile regressions. In *Handbook of Statistics* **30** 213–257. Elsevier, Amsterdam.

XIAO, Z. and KOENKER, R. (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *J. Amer. Statist. Assoc.* **104** 1696–1712. MR2750586 https://doi.org/10.1198/jasa.2009.tm09170

ZHANG, N. and ZHU, K. (2026). Supplement to “Quantiled conditional variance, skewness, and kurtosis by Cornish-Fisher expansion.” https://doi.org/10.1214/25-AOAS2122SUPP

ZHENG, Y., ZHU, Q., LI, G. and XIAO, Z. (2018). Hybrid quantile regression estimation for time series models with conditional heteroscedasticity. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 975–993. MR3874306 https://doi.org/10.1111/rssb.12277

ZHU, K. (2023). A new generalized exponentially weighted moving average quantile model and its statistical inference. *J. Econometrics* **237** 105510. MR4632721 https://doi.org/10.1016/j.jeconom.2023.105510

ZHU, Z., ZHANG, N. and ZHU, K. (2024). Big portfolio selection by graph-based conditional moments method. *J. Empir. Finance* **78** 101533.

ZHU, Z. and ZHU, K. (2025a). Alpha discovery in finance with distributional reinforcement learning. In *Proceedings of the* 42*nd International Conference on Machine Learning* (*ICML*), Vancouver, Cananda.

ZHU, Z. and ZHU, K. (2025b). Machine learning vast dynamic conditional covariance matrices: The spirit of “divide and conquer”. *Major Revision for Manag. Sci.*

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 148–172
https://doi.org/10.1214/25-AOAS2127

FEATURE AUGMENTATIONS FOR HIGH-DIMENSIONAL LEARNING: APPLICATIONS TO STOCK MARKET PREDICTION USING CHINESE NEWS DATA

BY XIAONAN ZHU[a], BINGYAN WANG[b] AND JIANQING FAN[c]

Department of Operations Research and Financial Engineering, Princeton University, [a]*xz8451@princeton.edu,* [b]*bingyanw@princeton.edu,* [c]*jqfan@princeton.edu*

High-dimensional measurements are often correlated, which motivates their approximation by factor models. This holds also true when features are engineered via low-dimensional interactions or kernel tricks. This often results in overparametrization and requires a fast dimensionality reduction. We propose a simple technique to enhance the performance of supervised learning algorithms by augmenting features with factors extracted from design matrices and their transformations. This is implemented by using the factors and idiosyncratic residuals which significantly weaken the correlations between input variables and hence increase the interpretability of learning algorithms and numerical stability. Extensive experiments on various algorithms and real-world data in diverse fields are carried out, among which we put special emphasis on the stock return prediction problem with Chinese financial news data due to the increasing interest in NLP problems in financial studies. We verify the capability of the proposed feature augmentation approach to boost overall prediction performance with the same algorithm. The approach bridges a gap in research that has been overlooked in previous studies, which focus either on collecting additional data or constructing more powerful algorithms, whereas our method lies in between these two directions using a simple PCA augmentation.

REFERENCES

AHN, S. C. and HORENSTEIN, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81** 1203–1227. MR3064065 https://doi.org/10.3982/ECTA8968

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

BAI, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71** 135–171. MR1956857 https://doi.org/10.1111/1468-0262.00392

BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259 https://doi.org/10.1111/1468-0262.00273

BALLI, H. O. and SØRENSEN, B. E. (2013). Interaction effects in econometrics. *Empir. Econ.* **45** 583–603.

CERQUETI, R., IOVANELLA, A., MATTERA, R. and STORANI, S. (2024). Improving the explainability of autoencoder factors for commodities through forecast-based Shapley values. *Sci. Rep.* **14** 19622.

CHEN, E., FAN, J. and ZHU, X. (2024). Factor augmented matrix regression. arXiv Preprint. Available at arXiv:2405.17744.

CHEN, R., YANG, D. and ZHANG, C.-H. (2022). Factor models for high-dimensional tensor time series. *J. Amer. Statist. Assoc.* **117** 94–116. MR4399070 https://doi.org/10.1080/01621459.2021.1912757

COCHRANE, J. H. and PIAZZESI, M. (2005). Bond risk premia. *Amer. Econ. Rev.* **95** 138–160.

FAMA, E. F. and FRENCH, K. R. (1992). The cross-section of expected stock returns. *J. Finance* **47** 427–465.

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. MR3494659 https://doi.org/10.1080/01621459.2015.1005212

FAN, J. and GU, Y. (2024). Factor augmented sparse throughput deep ReLU neural networks for high dimensional regression. *J. Amer. Statist. Assoc.* **119** 2680–2694. MR4833907 https://doi.org/10.1080/01621459.2023.2271605

Key words and phrases. Factor augmentations, principal components, feature interactions, kernel features, prediction.

FAN, J., GUO, J. and ZHENG, S. (2022). Estimating number of factors by adjusted eigenvalues thresholding. *J. Amer. Statist. Assoc.* **117** 852–861. MR4436317 https://doi.org/10.1080/01621459.2020.1825448

FAN, J., KE, Y. and WANG, K. (2020). Factor-adjusted regularized model selection. *J. Econometrics* **216** 71–85. MR4077382 https://doi.org/10.1016/j.jeconom.2020.01.006

FAN, J. and LIAO, Y. (2022). Learning latent factors from diversified projections and its applications to over-estimated and weak factors. *J. Amer. Statist. Assoc.* **117** 909–924. MR4436322 https://doi.org/10.1080/01621459.2020.1831927

FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x

GOLDSTEIN, I., SPATT, C. S. and YE, M. (2021). Big data in finance. *Rev. Financ. Stud.* **34** 3213–3225.

GRINSZTAJN, L., OYALLON, E. and VAROQUAUX, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Adv. Neural Inf. Process. Syst.* **35** 507–520.

HIGUERA, C., GARDINER, K. J. and CIOS, K. J. (2015). Self-organizing feature maps identify proteins critical to learning in a mouse model of down syndrome. *PLoS ONE* **10** e0129126.

HINTON, G. E., SRIVASTAVA, N., KRIZHEVSKY, A., SUTSKEVER, I. and SALAKHUTDINOV, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. arXiv Preprint. Available at arXiv:1207.0580.

KE, Z. T., KELLY, B. T. and XIU, D. (2019). Predicting returns with text data Technical Report National Bureau of Economic Research.

WANG, L., CHENG, Y., XIANG, A., ZHANG, J. and YANG, H. (2024). Application of natural language processing in financial risk detection. arXiv Preprint. Available at arXiv:2406.09765.

KRIZHEVSKY, A., HINTON, G. et al. (2009). Learning multiple layers of features from tiny images.

LAM, C. and YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** 694–726. MR2933663 https://doi.org/10.1214/12-AOS970

LECUN, Y. (1998). The MNIST database of handwritten digits. Available at http://yann.Lecun.Com/exdb/mnist/.

LOUGHRAN, T. and MCDONALD, B. (2016). Textual analysis in accounting and finance: A survey. *J. Acc. Res.* **54** 1187–1230.

MACHMUD, R., WIJAYA, A. et al. (2016). Behavior determinant based cervical cancer early detection with machine learning algorithm. *Adv. Sci. Lett.* **22** 3120–3123.

MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* (*N.S.*) **72** 507–536. MR0208649

MCCRACKEN, M. W. and NG, S. (2016). FRED-MD: A monthly database for macroeconomic research. *J. Bus. Econom. Statist.* **34** 574–589. MR3547997 https://doi.org/10.1080/07350015.2015.1086655

RITCHIE, H., MATHIEU, E., RODES-GUIRAO, L., APPEL, C., GIATTINO, C., ORTIZ-OSPINA, E., HASELL, J., MACDONALD, B., BELTEKIAN, D. et al. (2020). Coronavirus pandemic (COVID-19). Our World in Data. Available at https://ourworldindata.org/coronavirus.

RODRIGUES, F., MARKOU, I. and PEREIRA, F. C. (2019). Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach. *Inf. Fusion* **49** 120–129.

SHWARTZ-ZIV, R. and ARMON, A. (2022). Tabular data: Deep learning is not all you need. *Inf. Fusion* **81** 84–90.

STOCK, J. H. and WATSON, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *J. Bus. Econom. Statist.* **20** 147–162. MR1963257 https://doi.org/10.1198/073500102317351921

SUN, J. (2017). Jieba: Chinese text segmentation. Available at https://github.com/fxsjy/jieba. Accessed: April 2025.

TSAI, S.-C., LIN, S.-J., CHEN, P.-W., LUO, W.-Y., YEH, T.-H., WANG, H.-W., CHEN, C.-J. and TSAI, C.-H. (2009). EBV Zta protein induces the expression of interleukin-13, promoting the proliferation of EBV-infected B cells and lymphoblastoid cell lines. *Blood* **114** 109–118.

WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics*: *A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

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

WU, M.-Y., ZHANG, X.-F., DAI, D.-Q., OU-YANG, L., ZHU, Y. and YAN, H. (2016). Regularized logistic regression with network-based pairwise interaction for biomarker identification in breast cancer. *BMC Bioinform.* **17** 1–18.

XIAO, H., RASUL, K. and VOLLGRAF, R. (2017). Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms. arXiv Preprint. Available at arXiv:1708.07747.

XIU, D. and SHEN, Z. (2024). Deep autoencoders for nonlinear factor models: Theory and applications. Available at SSRN.

ZHANG, D., ZHANG, H., ZHOU, H., BAO, X., HUO, D., CHEN, R., CHENG, X., WU, M. and ZHANG, Q. (2021). Building interpretable interaction trees for deep nlp models. In *Proceedings of the AAAI Conference on Artificial Intelligence* **35** 14328–14337.

ZHANG, L., ZHOU, W. and WANG, H. (2022). Non-asymptotic properties of spectral decomposition of large Gram-type matrices and applications. *Bernoulli* **28** 1224–1249. MR4388936 https://doi.org/10.3150/21-bej1384

ZHOU, Y., FAN, J. and XUE, L. (2024). How much can machines learn finance from Chinese text data? *Manag. Sci.*

ZHOU, Y., XUE, L., SHI, Z., WU, L. and FAN, J. (2022). Measuring housing vitality from multi-source big data and machine learning. *J. Amer. Statist. Assoc.* **117** 1045–1059. MR4480687 https://doi.org/10.1080/01621459.2022.2096038

ZHU, X., WANG, B. and FAN, J. (2026). Supplement to "Feature augmentations for high-dimensional learning: Applications to stock market prediction using Chinese news data." https://doi.org/10.1214/25-AOAS2127SUPPA, https://doi.org/10.1214/25-AOAS2127SUPPB

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 173–196
https://doi.org/10.1214/26-AOAS2137

TEMPERATURE IN THE IBERIAN PENINSULA: COMMON TRENDS AND HETEROGENEITY

BY C. VLADIMIR RODRÍGUEZ-CABALLERO[1,a] AND ESTHER RUIZ[2,b]

[1] *Department of Statistics, ITAM,* [a]*vladimir.rodriguez@itam.mx*
[2] *Department of Statistics, Universidad Carlos III de Madrid,* [b]*ortega@est-econ.uc3m.es*

We propose a Multilevel Dynamic Factor Model (ML-DFM) to capture the common global and region-specific stochastic trends in monthly centre and log-range temperatures observed at 68 locations across the Iberian Peninsula from January 1930 to December 2020. The specification of common trends is based on the analysis of temperatures at each location using unobserved component models, which decompose temperatures into trend, seasonal, and transitory components. First, we show that the centre and log-range temperatures evolve independently. Second, we remove the seasonal component before analysing common trends. Third, we find that centre temperature trends are well approximated by a smooth, integrated random walk with a time-varying slope. In contrast, a stochastic level better captures the dynamics of the log-range. The ML-DFM is estimated using an EM algorithm extended here to accommodate nonstationary factors. We show that, although the commonality in centre-temperature trends is considerable, the regional components remain relevant, particularly at the log-range.

REFERENCES

BAI, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71** 135–171. MR1956857 https://doi.org/10.1111/1468-0262.00392

BANBURA, M., GIANNONE, D. and REICHLIN, L. (2011). Nowcasting. In *Oxford Handbook of Economic Forecasting* (M. P. Clements and D. F. Hendry, eds.) Oxford Univ. Press, Oxford.

BELLOCCA, G. P., GARRÓN, I., RODRÍGUEZ-CABALLERO, C. V. and RUIZ, E. (2025). FARS: Factor Augmented Regression Scenarios in R. Available at arXiv:2507.10679 [stat.CO].

BINELLI, C., LOVELESS, M. and SCHAFFNER, B. F. (2023). Explaining perceptions of climate change in the US. *Polit*. *Res*. *Q*. **76** 365–380.

BLOOMFIELD, P. (1992). Trends in global temperature. *Clim*. *Change* **21** 1–16.

BOGALO, J., PONCELA, P. and SENRA, E. (2024). Understanding fluctuations through multivariate circulant singular spectrum analysis. *Expert Syst*. *Appl*. 123827.

BREITUNG, J. and EICKMEIER, S. (2016). Analyzing international business and financial cycles using multilevel factor models: A comparison of alternative approaches. In *Dynamic Factor Models* **35** 177–214. Emerald Group Publishing Limited.

BUSETTI, F. and HARVEY, A. (2003). Seasonality tests. *J*. *Bus*. *Econom*. *Statist*. **21** 420–436. MR1997575 https://doi.org/10.1198/073500103288619061

CAMACHO, M., LOVCHA, Y. and PEREZ QUIROS, G. (2015). Can we use seasonally adjusted variables in dynamic factor models? *Stud*. *Nonlinear Dyn*. *Econom*. **19** 377–391. MR3355605 https://doi.org/10.1515/snde-2013-0096

CAMPBELL, S. D. and DIEBOLD, F. X. (2005). Weather forecasting for weather derivatives. *J*. *Amer*. *Statist*. *Assoc*. **100** 6–16. MR2166065 https://doi.org/10.1198/016214504000001051

CHANG, Y., KAUFMANN, R. K., KIM, C. S., MILLER, J. I., PARK, J. Y. and PARK, S. (2020). Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate. *J*. *Econometrics* **214** 274–294. MR4038232 https://doi.org/10.1016/j.jeconom.2019.05.014

CICCARELLI, M., KUIK, F. and HERNÁNDEZ, C. M. (2024). The asymmetric effects of temperature shocks on inflation in the largest euro area countries. *Eur*. *Econ*. *Rev*. **168** 104805.

COGGIN, T. D. (2012). Using econometric methods to test for trends in the HadCRUT3 global and hemispheric data. *Int*. *J*. *Climatol*. **32** 315–320.

Key words and phrases. Climate change, dynamic factor models, Kalman filter, state space models.

CORONA, F., PONCELA, P. and RUIZ, E. (2020). Estimating non-stationary common factors: Implications for risk sharing. *Comput. Econ.* **55** 37–60.

DELL, M., JONES, B. F. and OLKEN, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *J. Econ. Lit.* **52** 740–798.

DENG, Q. and FU, Z. (2019). Comparison of methods for extracting annual cycle with changing amplitude in climate series. *Clim. Dyn.* **52** 5059–5070.

DESMET, K. and ROSSI-HANSBERG, E. (2024). Climate change economics over time and space. *Ann. Rev. Econ.* **16** 271–304.

DIEBOLD, F. X. and RUDEBUSCH, G. D. (2022a). On the evolution of U.S. temperature dynamics. In *Essays in Honor of Hashem Pesaran: Prediction and Macro Modeling*, *Advances in Econometrics*, 43*A* (A. Chudik, C. Hsiao and A. Timmermann, eds.) Emerald Publishing Limited.

DIEBOLD, F. X. and RUDEBUSCH, G. D. (2022b). Probability assessments of an ice-free Arctic: Comparing statistical and climate model projections. *J. Econometrics* **231** 520–534. MR4500727 https://doi.org/10.1016/j.jeconom.2020.12.007

DOZ, C., GIANNONE, D. and REICHLIN, L. (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models. *Rev. Econ. Stat.* **94** 1014–1024.

DUCA, V. EL. A., FONSECA, T. CO. and OLIVEIRA, F. LC. (2023). An overview of non-Gaussian state-space models for wind speed data. *Energy* **266** 126436.

DUPUIS, D. J. (2012). Modeling waves of extreme temperature: The changing tails of four cities. *J. Amer. Statist. Assoc.* **107** 24–39. MR2949339 https://doi.org/10.1080/01621459.2011.643732

DUPUIS, D. J. (2014). A model for nighttime minimum temperatures. *J. Climate* **27** 7207–7229.

ERGEMEN, Y. E. and RODRÍGUEZ-CABALLERO, C. V. (2023). Estimation of a dynamic multilevel factor model with possible long-range dependence. *Int. J. Forecast.* **39** 405–430.

ESTRADA, F., KIM, D. and PERRON, P. (2021). Spatial variations in the warming trend and the transition to more severe weather in midlatitudes. *Sci. Rep.* **11**.

ESTRADA, F. and PERRON, P. (2017). Extracting and analyzing the warming trend in global and hemispheric temperatures. *J. Time Series Anal.* **38** 711–732. MR3689442 https://doi.org/10.1111/jtsa.12246

EVERITT, B. S., LANDAU, S., LEESE, M. and STAHL, D. (2011). *Cluster Analysis*, 5th ed. Wiley.

ESTRADA, F. and PERRON, P. (2021). Disentangling the trend in the warming of urban areas into global and local factors. *Ann. N.Y. Acad. Sci.* **1504** 230–246.

GADEA RIVAS, M. D. and GONZALO, J. (2020). Trends in distributional characteristics: Existence of global warming. *J. Econometrics* **214** 153–174. MR4038227 https://doi.org/10.1016/j.jeconom.2019.05.009

GONZALEZ-RIVERA, G., LUO, Y. and RUIZ, E. (2020). Prediction regions for interval-valued time series. *J. Appl. Econometrics* **35** 373–390. MR4114488 https://doi.org/10.1002/jae.2754

GOOD, S., CORLETT, G., REMEDIOS, J., NOYES, E. and LLEWELLYN-JONES, D. (2007). The global trend in sea surface temperature from 20 years of advanced very high resolution radiometer data. *J. Climate* **20** 1255–1264.

GOSPODINOV, N., LOPEZ GAFFENY, I. and NG, S. (2025). The economic impact of low- and high-frequency temperature changes. Available at arXiv:2505.08950v1 [econ.GN].

HANNAN, E. J. and DEISTLER, M. (1988). *The Statistical Theory of Linear Systems*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0940698

HARRIS, I., OSBORN, T. J., JONES, P. and LISTER, D. (2020). Version 4 of the CRUTS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* **7**.

HARVEY, A. C. (1989). *Forecasting*, *Structural Time Series Models and the Kalman Filter*. Cambridge Univ. Press, Cambridge.

HARVEY, A. C. (2001). Testing in unobserved components models. *J. Forecast.* **20** 1–19.

HARVEY, A. C. (2016). *Trend Analysis*, *Wiley StatsRef*: *Statistics Reference Online*.

HELSKE, J. (2017). KFAS: Exponential family state space models in R. *J. Stat. Softw.* **78**.

HILLEBRAND, E. and PROIETTI, T. (2017). Phase changes and seasonal warming in early instrumental temperature records. *J. Climate* **30** 6795–6821.

HINDRAYANTO, I., ASTON, J. A. D., KOOPMAN, S. J. and OOMS, M. (2013). Modeling trigonometric seasonal components for monthly economic time series. *Appl. Econ.* **45** 3024–3034.

HOLT, M. T. and TERÄSVIRTA, T. (2020). Global hemispheric temperatures and co-shifting: A vector shifting-mean autoregressive analysis. *J. Econometrics* **214** 198–215. MR4038229 https://doi.org/10.1016/j.jeconom.2019.05.011

IPCC (2023). AR6 Synthesis Report: Climate Change 2022. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. [Core Writing Team, Aldunce, P. et al.]. IPCC, Geneva, Switzerland.

KATZ, R. W. and BROWN, B. G. (1992). Extreme events in a changing climate: Variability is more important than averages. *Clim. Change* **21** 289–302.

KAUFMANN, R. K., KAUPPI, H., MANN, M. L. and STOCK, J. H. (2013). Does temperature contain a stochastic trend: Linking statistical results to physical mechanisms. *Clim. Change* **118** 729–743.

KAUFMANN, R. K., KAUPPI, H. and STOCK, J. H. (2010). Does temperature contain a stochastic trend? Evaluating conflicting statistical results. *Clim. Change* **101** 395–405.

KAUFMANN, R. K., MANN, M. L., GOPAL, S., LIEDERMAN, J. A., HOWE, P. D., PRETIS, F., TANG, X. and GILMORE, M. (2017). Spatial heterogeneity of climate change as an experimental basis for skepticism. *Proc. Natl. Acad. Sci. PNAS* **114** 67–71.

KEW, S. F., PHILIP, S. Y., JAN VAN OLDENBORGH, G., VAN DER SCHRIER, G., OTTO, F. E. and VAUTARD, R. (2019). The exceptional summer heat wave in Southern Europe 2017. *Bull. Amer. Meteorol. Soc.* **100** S49–S53.

MAROTTA, F. and MUMTAZ, H. (2023). Vulnerability to climate change: Evidence from a dynamic factor model. Oxford Smith School of Enterprise and the Environment Working Paper no. 23-06.

MEDHAUG, I., STOLPE, M. B., FISCHER, E. M. and KNUTTI, R. (2017). Reconciling controversies about the 'global warming hiatus'. *Nature* **545** 41–47.

MENG, X. and TAYLOR, J. W. (2022). Comparing probabilistic forecasts of the daily minimum and maximum temperature. *Int. J. Forecast.* **38** 267–281.

MILLER, J. I. and NAM, K. (2020). Dating hiatuses: A statistical model of the recent slow-down in global warming and the next one. *Earth Syst. Dyn.* **11** 1123–1132.

MILLER, S., CHUA, K., COGGINGS, J. and MOHTADI, H. (2021). Heat waves, climate change, and economic output. *J. Eur. Econ. Assoc.* **19** 2658–2694.

MITCHELL, T. D. and JONES, P. D. (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.* **25** 693–712.

MUDELSEE, M. (2019). Trend analysis of climate time series: A review. *Earth-Sci. Rev.* **190** 310–322.

NYBLOM, J. and HARVEY, A. C. (2000). Testing against smooth stochastic trends. *J. Appl. Econometrics* **16** 415–429.

PAPANASTASSIOU, D. (2006). Computing the covariance matrix of QML estimators for a state space model. *Statist. Probab. Lett.* **76** 1001–1006. MR2269334 https://doi.org/10.1016/j.spl.2005.11.002

PEZZULLI, S., STEPHENSON, D. and HANNACHI, A. (2005). The variability of seasonality. *J. Climate* **18** 71–88.

PRETIS, F. and HENDRY, D. F. (2013). Comment on "Polynomial cointegration tests of anthropogenic impact on global warming" by Beenstock et al. (2012)-some hazards in econometric modelling of climate change. *Earth Syst. Dyn.* **4** 375–384.

PRETIS, F., MANN, M. L. and KAUFMANN, R. K. (2015). Testing competing models of the temperature hiatus: Assessing the effects of conditioning variables and temproal uncertainties through sample-wide break detection. *Clim. Change* **131** 705–718.

PROIETTI, T. and HILLEBRAND, E. (2015). Seasonal changes in central England temperatures. *J. Roy. Statist. Soc., Ser. A, Statist. Soc.* **180** 769–791.

RODRÍGUEZ-CABALLERO, C. V. and RUIZ, E. (2026). Supplement to "Temperature in the Iberian Peninsula: Common trends and heterogeneity." https://doi.org/10.1214/26-AOAS2137SUPPA, https://doi.org/10.1214/26-AOAS2137SUPPB

SCHLEMM, E. and STELZER, R. (2012). Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Lévy-driven CARMA processes. *Electron. J. Stat.* **6** 2185–2234. MR3020261 https://doi.org/10.1214/12-EJS743

SCHMIDT, G. A., SHINDELL, D. T. and TSIGARIDIS, K. (2014). Reconciling warming trends. *Nat. Geosci.* **7** 158–160.

SHUMWAY, R. H. and STOFFER, D. S. (2016). *Time Series Analysis and Its Applications*, 4th ed. Springer, Berlin.

STERN, D. I. and KAUFMANN, R. K. (2000). Detecting a global warming signal in hemispheric temperature series: A structural time series analysis. *Clim. Change* **47** 411–438.

VISSER, H. (2004). Estimation and detection of flexible trends. *Atmos. Environ.* **38** 4135–4145.

VISSER, H. and MOLENAAR, J. (1995). Trend estimation and regression analysis in climatological time series: An application of structural time series models and the Kalman filter. *J. Climate* **8** 969–979.

VOSE, R. S., EASTERLING, D. R. and GLEASON, B. (2005). Maximum and minimum temperature trends for the globe: An update through 2004. *Geophys. Res. Lett.* **32** L23822.

WIJNGAARD, J. B., KLEIN TAUK, A. M. G. and KÖNNEN, G. P. (2003). Homogeneity of 20th century European daily temperature and precipitations series. *Int. J. Climatol.* **23** 679–692.

WOODWARD, W. H. and GRAY, H. L. (1993). Global warming and the problem of testing for trend in time series data. *J. Climate* **6** 953–962.

ZAVAL, L., KEENAN, E. A., JOHNSON, E. J. and WEBER, E. U. (2014). How warm days increase belief in global warming. *Nat. Clim. Change* **4** 143–147.

ZHENG, X. and BASHER, R. E. (1999). Structural time series models and trend detection in global and regional temperature series. *J. Climate* **12** 2347–2358.

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 197–214
https://doi.org/10.1214/26-AOAS2141

REGIONALIZATION OF CHINA'S PM$_{2.5}$: A ROBUST FUNCTIONAL SPATIAL CLUSTERING WITH ANGULAR DEPTH

BY TINGYIN WANG[1,a], XUEQIN WANG[1,b], XIAOBO GUO[2,c] AND HEPING ZHANG[3,d]

[1]*Department of Statistics and Finance, School of Management, University of Science and Technology of China,* [a]*christinawang666@mail.ustc.edu.cn,* [b]*wangxq20@mail.ustc.edu.cn*

[2]*Department of Statistical Science, School of Mathematics, Sun Yat-Sen University,* [c]*guoxb3@mail.sysu.edu.cn*

[3]*Department of Biostatistics, School of Public Health, Yale University,* [d]*heping.zhang@yale.edu*

Particulate matter with aerodynamic diameters smaller than 2.5 μm (PM$_{2.5}$) exhibits substantial spatial variation across China, characterized by heterogeneous patterns at the national scale and relative homogeneity within smaller regions. Analyzing these patterns is particularly challenging due to strong spatial similarity among neighboring sites and the presence of outliers in the data. To address these challenges, we propose a robust functional spatial clustering framework built upon the concept of angular depth, which provides a robust centrality measure for functional data with desirable theoretical properties in infinite-dimensional spaces. Leveraging angular depth, our method effectively accommodates outliers and incorporates spatial information to produce stable and interpretable clustering results. Applying the proposed framework to a national PM$_{2.5}$ dataset, we identify 10 distinct regions with well-defined boundaries and internally coherent pollution patterns. The resulting clusters offer valuable insights for policymakers, providing a scientific basis for designing targeted emission-control strategies and fostering regional cooperation in air quality management.

REFERENCES

BELL, M. L., EBISU, K., PENG, R. D. et al. (2007). Seasonal and regional short-term effects of fine particles on hospital admissions in 202 US counties, 1999–2005. *Amer. J. Epidemiol.* **168** 1301–1310.

BOUVEYRON, C., CÔME, E. and JACQUES, J. (2015). The discriminative functional mixture model for a comparative analysis of bike sharing systems. *Ann. Appl. Stat.* **9** 1726–1760. MR3456352 https://doi.org/10.1214/15-AOAS861

CHEN, L.-J., HO, Y.-H., HSIEH, H.-H., HUANG, S.-T., LEE, H.-C. and MAHAJAN, S. (2017). Adf: An anomaly detection framework for large-scale PM2.5 sensing systems. *IEEE Internet Things J.* **5** 559–570.

CHENG, J., SU, J., CUI, T., LI, X., DONG, X., SUN, F., YANG, Y., TONG, D., ZHENG, Y. et al. (2019). Dominant role of emission reduction in PM2.5 air quality improvement in Beijing during 2013–2017: A model-based decomposition analysis. *Atmos. Chem. Phys.* **19** 6125–6146.

CHU, H.-J., HUANG, B. and LIN, C.-Y. (2015). Modeling the spatio-temporal heterogeneity in the PM10–PM2.5 relationship. *Atmos. Environ.* **102** 176–182.

CUESTA-ALBERTOS, J. A. and NIETO-REYES, A. (2008). The random Tukey depth. *Comput. Statist. Data Anal.* **52** 4979–4988. MR2526207 https://doi.org/10.1016/j.csda.2008.04.021

CUEVAS, A., FEBRERO, M. and FRAIMAN, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comput. Statist.* **22** 481–496. MR2336349 https://doi.org/10.1007/s00180-007-0053-0

CUI, M. et al. (2020). Introduction to the k-means clustering algorithm based on the elbow method. *J. Account. Audit. Financ.* **1** 5–8.

DE BOOR, C. (1978). *A Practical Guide to Splines. Applied Mathematical Sciences* **27**. Springer, New York. MR0507062

FANG, C., TAN, X., ZHONG, Y. and WANG, J. (2021). Research on the temporal and spatial characteristics of air pollutants in Sichuan Basin. *Atmosphere* **12** 1504.

FEBRERO, M., GALEANO, P. and GONZÁLEZ-MANTEIGA, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal NO$_x$ levels. *Environmetrics* **19** 331–345. MR2440036 https://doi.org/10.1002/env.878

Key words and phrases. Functional depth, robust clustering, environmental policies.

HOPKE, P. K., DAI, Q., LI, L. and FENG, Y. (2020). Global review of recent source apportionments for airborne particulate matter. *Sci*. *Total Environ*. **740** 140091.

HU, G., GENG, J., XUE, Y. and SANG, H. (2023). Bayesian spatial homogeneity pursuit of functional data: An application to the U.S. income distribution. *Bayesian Anal*. **18** 579–605. MR4578065 https://doi.org/10.1214/22-ba1320

HUANG, R.-J., ZHANG, Y., BOZZETTI, C., HO, K.-F., CAO, J.-J., HAN, Y., DAELLENBACH, K. R., SLOWIK, J. G., PLATT, S. M. et al. (2014). High secondary aerosol contribution to particulate pollution during haze events in China. *Nature* **514** 218–222.

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

HUBERT, M., ROUSSEEUW, P. J. and SEGAERT, P. (2015). Multivariate functional outlier detection. *Stat*. *Methods Appl*. **24** 177–202. MR3376852 https://doi.org/10.1007/s10260-015-0297-8

JIANG, L., HE, S. and ZHOU, H. (2020). Spatio-temporal characteristics and convergence trends of PM2.5 pollution: A case study of cities of air pollution transmission channel in Beijing–Tianjin–Hebei region, China. *J*. *Clean*. *Prod*. **256** 120631.

KELLY, F. J. and FUSSELL, J. C. (2012). Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. *Atmos*. *Environ*. **60** 504–526.

KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020). Robust multivariate nonparametric tests via projection averaging. *Ann*. *Statist*. **48** 3417–3441. MR4185814 https://doi.org/10.1214/19-AOS1936

KIMMEL, R. and SETHIAN, J. A. (1998). Computing geodesic paths on manifolds. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **95** 8431–8435. MR1639135 https://doi.org/10.1073/pnas.95.15.8431

KNEIP, A. and LIEBL, D. (2020). On the optimal reconstruction of partially observed functional data. *Ann*. *Statist*. **48** 1692–1717. MR4124340 https://doi.org/10.1214/19-AOS1864

KRAUS, D. (2015). Components and completion of partially observed functional data. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **77** 777–801. MR3382597 https://doi.org/10.1111/rssb.12087

LI, J., TAN, J. and WANG, X. (2025). Sparse equation matching: a derivative-free learning for general-order dynamical systems. arXiv preprint. Available at arXiv:2507.20072.

LIANG, D., ZHANG, H., CHANG, X. and HUANG, H. (2021). Modeling and regionalization of China's $\mathrm{PM}_{2.5}$ using spatial-functional mixture models. *J*. *Amer*. *Statist*. *Assoc*. **116** 116–132. MR4227679 https://doi.org/10.1080/01621459.2020.1764363

LÓPEZ-PINTADO, S. and ROMO, J. (2009). On the concept of depth for functional data. *J*. *Amer*. *Statist*. *Assoc*. **104** 718–734. MR2541590 https://doi.org/10.1198/jasa.2009.0108

LÓPEZ-PINTADO, S. and ROMO, J. (2011). A half-region depth for functional data. *Comput*. *Statist*. *Data Anal*. **55** 1679–1695. MR2748671 https://doi.org/10.1016/j.csda.2010.10.024

LUO, F., TAN, J., ZHANG, D., HUANG, H. and SHEN, Y. (2025). Functional clustering for longitudinal associations between social determinants of health and stroke mortality in the U.S. *Ann*. *Appl*. *Stat*. **19** 798–820. MR4888133 https://doi.org/10.1214/24-aoas1989

RAMSAY, J. O. and SILVERMAN, B. (1997). Principal differential analysis. In *Functional Data Analysis* 239–256. Springer, Berlin.

REN, C., WU, L., ZHANG, Y., LI, J., CHAI, M., XIANG, C. et al. (2016). Analyze to the seasonal differences of transport pathways and potential source-zones of Beijing urban PM2.5. *China Environ*. *Sci*. **36** 2591–2598.

SHEN, L., WEN, J., ZHANG, Y., ULLAH, S., CHENG, J. and MENG, X. (2022). Changes in population exposure to extreme precipitation in the Yangtze River Delta, China. *Clim*. *Serv*. **27** 100317.

SONG, F., LIANG, D. and ZOU, C. (2025). Change-points detection and support recovery for spatially indexed functional data. arXiv preprint. Available at arXiv:2506.07206.

SUN, Y. and GENTON, M. G. (2011). Functional boxplots. *J*. *Comput*. *Graph*. *Statist*. **20** 316–334. MR2847798 https://doi.org/10.1198/jcgs.2011.09224

TAN, J., GE, Y., MARTINEZ, L., SUN, J., LI, C., WESTBROOK, A., CHEN, E., PAN, J., LI, Y. et al. (2022). Transmission roles of symptomatic and asymptomatic covid-19 cases: A modelling study. *Epidemiol*. *Infect*. **150** e171.

TAN, J., LIANG, D., GUAN, Y. and HUANG, H. (2024a). Graphical principal component analysis of multivariate functional time series. *J*. *Amer*. *Statist*. *Assoc*. **119** 3073–3085. MR4833938 https://doi.org/10.1080/01621459.2024.2302198

TAN, J., SHEN, Y., GE, Y., MARTINEZ, L. and HUANG, H. (2023). Age-related model for estimating the symptomatic and asymptomatic transmissibility of COVID-19 patients. *Biometrics* **79** 2525–2536. MR4644013 https://doi.org/10.1111/biom.13814

TAN, J., SHI, P. and ZHANG, A. R. (2024). Functional singular value decomposition. arXiv preprint. Available at arXiv:2410.03619.

TAN, J., ZHANG, G., WANG, X., HUANG, H. and YAO, F. (2024b). Green's matching: An efficient approach to parameter estimation in complex dynamic systems. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **86** 1266–1285. MR4896637 https://doi.org/10.1093/jrsssb/qkae031

TIAN, T., TAN, J., LUO, W., JIANG, Y., CHEN, M., YANG, S., WEN, C., PAN, W. and WANG, X. (2021). The effects of stringent and mild interventions for coronavirus pandemic. *J. Amer. Statist. Assoc.* **116** 481–491. MR4269997 https://doi.org/10.1080/01621459.2021.1897015

TSAI, H.-H., YUAN, C.-S., HUNG, C.-H., LIN, C. et al. (2011). Physicochemical properties of PM2.5 and PM2.5–10 at inland and offshore sites over southeastern coastal region of Taiwan Strait. *Aerosol Air Qual. Res.* **11** 664–678.

WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.

WANG, P., CAO, J.-J., SHEN, Z.-X., HAN, Y.-M., LEE, S.-C., HUANG, Y., ZHU, C.-S., WANG, Q.-Y., XU, H.-M. et al. (2015). Spatial and seasonal variations of PM2.5 mass and species during 2010 in xi'an, China. *Sci. Total Environ.* **508** 477–487.

WANG, S., LIU, X., YANG, X., ZOU, B. and WANG, J. (2018). Spatial variations of PM2.5 in Chinese cities for the joint impacts of human activities and natural conditions: A global and local regression perspective. *J. Clean. Prod.* **203** 143–152.

WANG, T., WANG, X., GUO, X. and ZHANG, H. (2026). Supplement to "Regionalization of China's $PM_{2.5}$: a Robust Functional Spatial Clustering with Angular Depth" https://doi.org/10.1214/26-AOAS2141SUPPA, https://doi.org/10.1214/26-AOAS2141SUPPB

WANG, X., WANG, Q., DUAN, Y. and HUANG, K. (2021). Complex network analysis of PM2.5 transport in the Yangtze River Delta region, China. *Stoch. Environ. Res. Risk Assess.* **35** 2645–2658.

XU, F. and BEARD, K. (2021). A comparison of prospective space-time scan statistics and spatiotemporal event sequence based clustering for covid-19 surveillance. *PLoS ONE* **16** e0252990.

YANG, L., QIN, C., LI, K., DENG, C. and LIU, Y. (2023). Quantifying the spatiotemporal heterogeneity of PM2.5 pollution and its determinants in 273 cities in China. *Int. J. Environ. Res. Public Health* **20** 1183.

YAO, L., YANG, L., YUAN, Q., YAN, C., DONG, C., MENG, C., SUI, X., YANG, F., LU, Y. et al. (2016). Sources apportionment of pm 2.5 in a background site in the North China plain. *Sci. Total Environ.* **541** 590–598.

ZHANG, B., SANG, H., LUO, Z. T. and HUANG, H. (2023). Bayesian clustering of spatial functional data with application to a human mobility study during COVID-19. *Ann. Appl. Stat.* **17** 583–605. MR4539045 https://doi.org/10.1214/22-aoas1643

ZHANG, Y., ZHENG, W., FANG, H. and XIA, J. (2022). Clean heating in northern China: Regional investigations and roadmap studies for urban area towards 2050. *J. Clean. Prod.* **334** 130233.

ZHOU, X., ZHANG, T., LI, Z., TAO, Y., WANG, F., ZHANG, X., XU, C., MA, S. and HUANG, J. (2018). Particulate and gaseous pollutants in a petrochemical industrialized valley city, western China during 2013–2016. *Environ. Sci. Pollut. Res. Int.* **25** 15174–15190.

ZOU, B.-B., HUANG, X.-F., ZHANG, B., DAI, J., ZENG, L.-W., FENG, N. and HE, L.-Y. (2017). Source apportionment of pm2.5 pollution in an industrial city in southern China. *Atmos. Pollut. Res.* **8** 1193–1202.

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 215–237
https://doi.org/10.1214/26-AOAS2142

ENVIRONMENTAL RISK ASSESSMENT VIA NONHOMOGENEOUS HIDDEN SEMI-MARKOV MODELS WITH PENALIZED VECTOR AUTOREGRESSION

BY MARCO MINGIONE[1,a], PIERFRANCESCO ALAIMO DI LORO[2,b], FRANCESCO LAGONA[3,c] AND ANTONELLO MARUOTTI[4,5,2,d]

[1]*Department of Sports, Human and Health Sciences, University of Rome "Foro Italico",* [a]*marco.mingione@uniroma4.it*
[2]*Department Law, Economics, Politics and Modern Languages, LUMSA University,* [b]*p.alaimodiloro@lumsa.it*
[3]*Department of Political Sciences, Roma Tre University,* [c]*francesco.lagona@uniroma3.it*
[4]*Departement of Public Health and Epidemiology, Khalifa University,* [d]*antonello.maruotti@ku.ac.ae*
[5]*Center for Biotechnology, Khalifa University*

Motivated by the study of pollution trends in the city of Bergen, we introduce a flexible statistical framework for modeling multivariate air pollution data via a nonhomogeneous hidden semi-Markov vector autoregression. The hidden process captures unobserved environmental conditions, while the vector autoregressive structure accounts for temporal autocorrelation and cross-pollutant dependencies. The model further allows time-varying environmental conditions to influence both the average levels of pollutant concentrations and the duration of different transient states. Parameters are estimated via maximum likelihood using a tailored expectation-maximization (EM) algorithm, integrated with state-specific ℓ_1 regularization to control overfitting and automatically select relevant temporal lags. The proposal is tested on simulated data under different scenarios and then applied to daily concentrations of nitrogens and particulate matter recorded in an urban area. Environmental risk is assessed by a Shapley value-based decomposition that attributes marginal risk contributions. This approach offers a comprehensive framework for multivariate environmental risk modeling, enabling better identification of high-pollution episodes and informing policy interventions.

REFERENCES

ADRIAN, T. and BRUNNERMEIER, M. K. (2016). CoVaR. *Amer. Econ. Rev.* **106** 1705–1741.

ANDERSSON, A. (2021). Mechanisms for log normal concentration distributions in the environment. *Sci. Rep.* **11** 16418.

BARAGAÑO, D., RATIÉ, G., SIERRA, C., CHRASTNỲ, V., KOMÁREK, M. and GALLEGO, J. (2022). Multiple pollution sources unravelled by environmental forensics techniques and multivariate statistics. *J. Hazard. Mater.* **424** 127413.

BARBU, V. S. and LIMNIOS, N. (2009). *Semi-Markov Chains and Hidden Semi-Markov Models Toward Applications: Their Use in Reliability and DNA Analysis* **191**. Springer, Berlin.

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

BERGEN KOMMUNE (2023). Årsrapport luftkvalitet i Bergen 2023 Technical Report Bergen Kommune.

BERNARDI, M., MARUOTTI, A. and PETRELLA, L. (2017). Multiple risk measures for multivariate dynamic heavy–tailed models. *J. Empir. Finance* **43** 1–32.

BIERNACKI, C., CELEUX, G. and GOVAERT, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 719–725.

BOAZ, R. M., LAWSON, A. B. and PEARCE, J. L. (2019). Multivariate air pollution prediction modeling with partial missingness. *Environmetrics* **30** e2592. MR4021440 https://doi.org/10.1002/env.2592

BOUVEYRON, C., JACQUES, J., SCHMUTZ, A., SIMÕES, F. and BOTTINI, S. (2022). Co-clustering of multivariate functional data for the analysis of air pollution in the south of France. *Ann. Appl. Stat.* **16** 1400–1422. MR4455886 https://doi.org/10.1214/21-aoas1547

BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer Series in Statistics*. Springer, Heidelberg. MR2807761 https://doi.org/10.1007/978-3-642-20192-9

Key words and phrases. HSMM, dynamic mixture, air pollution, penalized VAR, risk measure.

CAO, C. (2024). Integration of ten years of daily weather, traffic, and air pollution data from Norway's six largest cities. *Sci. Data* **11** 744.

CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. *Springer Series in Statistics*. Springer, New York. MR2159833

CHATTERJEE, A. and LAHIRI, S. N. (2011). Bootstrapping lasso estimators. *J. Amer. Statist. Assoc.* **106** 608–625. MR2847974 https://doi.org/10.1198/jasa.2011.tm10159

EFRON, B. (2000). The bootstrap and modern statistics. *J. Amer. Statist. Assoc.* **95** 1293–1296. MR1825279 https://doi.org/10.2307/2669773

FINAZZI, F., SCOTT, E. M. and FASSÒ, A. (2013). A model-based framework for air quality indices and population risk evaluation, with an application to the analysis of Scottish air quality data. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **62** 287–308. MR3045878 https://doi.org/10.1111/rssc.12001

FONG, P. W., LI, W. K., YAU, C. W. and WONG, C. S. (2007). On a mixture vector autoregressive model. *Canad. J. Statist.* **35** 135–150. MR2345379 https://doi.org/10.1002/cjs.5550350112

GENG, G., ZHANG, Q., TONG, D., LI, M., ZHENG, Y., WANG, S. and HE, K. (2017). Chemical composition of ambient $PM_{2.5}$ over China and relationship to precursor emissions during 2005–2012. *Atmos. Chem. Phys.* **17** 9187–9203.

HADJ-AMAR, B., JEWSON, J. and VANNUCCI, M. (2024). Bayesian sparse vector autoregressive switching models with application to human gesture phase segmentation. *Ann. Appl. Stat.* **18** 2511–2531. MR4782500 https://doi.org/10.1214/24-aoas1892

HOSKOVEC, L., KOSLOVSKY, M. D., KOEHLER, K., GOOD, N., PEEL, J. L., VOLCKENS, J. and WILSON, A. (2023). Infinite hidden Markov models for multiple multivariate time series with missing data. *Biometrics* **79** 2592–2604. MR4644018 https://doi.org/10.1111/biom.13715

KHALILI, A. and CHEN, J. (2007). Variable selection in finite mixture of regression models. *J. Amer. Statist. Assoc.* **102** 1025–1038. MR2411662 https://doi.org/10.1198/016214507000000590

KOSLIK, J.-O. (2025). Hidden semi-Markov models with inhomogeneous state dwell-time distributions. *Comput. Statist. Data Anal.* **209** 108171. MR4880548 https://doi.org/10.1016/j.csda.2025.108171

LAGONA, F. and MINGIONE, M. (2025). Nonhomogeneous hidden semi-Markov models for toroidal data. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **74** 142–166. MR4849420 https://doi.org/10.1093/jrsssc/qlae049

LI, S. (2020). Debiasing the debiased Lasso with bootstrap. *Electron. J. Stat.* **14** 2298–2337. MR4112726 https://doi.org/10.1214/20-EJS1713

LIANG, D., ZHANG, H., CHANG, X. and HUANG, H. (2021). Modeling and regionalization of China's $PM_{2.5}$ using spatial-functional mixture models. *J. Amer. Statist. Assoc.* **116** 116–132. MR4227679 https://doi.org/10.1080/01621459.2020.1764363

LIAO, K., PARK, E. S., ZHANG, J., CHENG, L., JI, D., YING, Q. and YU, J. Z. (2021). A multiple linear regression model with multiplicative log-normal error term for atmospheric concentration data. *Sci. Total Environ.* **767** 144282.

MARTINEZ-ZARZOSO, I. and MARUOTTI, A. (2013). The environmental Kuznets curve: Functional form, time-varying heterogeneity and outliers in a panel setting. *Environmetrics* **24** 461–475. MR3137747 https://doi.org/10.1002/env.2232

MARUOTTI, A. and ALAIMO DI LORO, P. (2023). CO_2 emissions and growth: A bivariate bidimensional mean-variance random effects model. *Environmetrics* **34** e2793. MR4614191 https://doi.org/10.1002/env.2793

MARUOTTI, A., BULLA, J., LAGONA, F., PICONE, M. and MARTELLA, F. (2017). Dynamic mixtures of factor analyzers to characterize multivariate air pollutant exposures. *Ann. Appl. Stat.* **11** 1617–1648. MR3709572 https://doi.org/10.1214/17-AOAS1049

MERLO, L., MARUOTTI, A., PETRELLA, L. and PUNZO, A. (2022). Quantile hidden semi-Markov models for multivariate time series. *Stat. Comput.* **32** 61. MR4466973 https://doi.org/10.1007/s11222-022-10130-1

MINGIONE, M., ALAIMO DI LORO, P., LAGONA, F. and MARUOTTI, A. (2026). Supplement to "Environmental risk assessment via nonhomogeneous hidden semi-Markov models with penalized vector autoregression." https://doi.org/10.1214/26-AOAS2142SUPPA, https://doi.org/10.1214/26-AOAS2142SUPPB

O'CONNELL, J. and HØJSGAARD, S. (2011). Hidden semi Markov models for multiple observation sequences: The mhsmm package for R. *J. Stat. Softw.* **39** 1–22.

OTT, W. R. (1990). A physical explanation of the lognormality of pollutant concentrations. *J. Air Waste Manage. Assoc.* **40** 1378–1383.

OUYANG, W., GUO, B., CAI, G., LI, Q., HAN, S., LIU, B. and LIU, X. (2015). The washing effect of precipitation on particulate matter and the pollution dynamics of rainwater in downtown Beijing. *Sci. Total Environ.* **505** 306–314.

RICCIOTTI, L., PICONE, M., POLLICE, A. and MARUOTTI, A. (2025). A zero-inflated hidden semi-Markov model with covariate-dependent sojourn parameters for analysing marine data in the Venice lagoon. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **74** 506–529. MR4896628 https://doi.org/10.1093/jrsssc/qlae065

RUIZ-SUAREZ, S., LEOS-BARAJAS, V. and MORALES, J. M. (2022). Hidden Markov and semi-Markov models: When and why are these models useful for classifying states in time series data? *J. Agric. Biol. Environ. Stat.* **27** 339–363. MR4416787 https://doi.org/10.1007/s13253-021-00483-x

SHAN, X., CASEY, J. A., SHEARSTON, J. A. and HENNEMAN, L. R. (2024). Methods for quantifying source-specific air pollution exposure to serve epidemiology, risk assessment, and environmental justice. *GeoHealth* **8** e2024GH001188.

STÄDLER, N., BÜHLMANN, P. and VAN DE GEER, S. (2010). ℓ_1-penalization for mixture regression models. *TEST* **19** 209–256. MR2677722 https://doi.org/10.1007/s11749-010-0197-z

STÄDLER, N. and MUKHERJEE, S. (2013). Penalized estimation in high-dimensional hidden Markov models with state-specific graphical models. *Ann. Appl. Stat.* **7** 2157–2179. MR3161717 https://doi.org/10.1214/13-AOAS662

TAN, L., CHIONG, K. X. and MOON, H. R. (2021). Estimation of high-dimensional seemingly unrelated regression models. *Econometric Rev.* **40** 830–851. MR4296872 https://doi.org/10.1080/07474938.2021.1889195

TAVELLA, R. A., GALEAO DA ROSA MORAES, N., MACIEL AICK, C. D., RAMIRES, P. F., PEREIRA, N., SOARES, A. G. and DA SILVA JÚNIOR, F. M. R. (2023). Weekend effect of air pollutants in small and medium-sized cities: The role of policies stringency to COVID-19 containment. *Atmos. Pollut. Res.* **14** 101662.

THUNIS, P., CLAPPIER, A., BEEKMANN, M., PUTAUD, J. P., CUVELIER, C., MADRAZO, J. and DE MEIJ, A. (2021). Non-linear response of $\mathrm{PM}_{2.5}$ to changes in NO_x and NH_3 emissions in the Po basin (Italy): Consequences for air quality plans. *Atmos. Chem. Phys.* **21** 9309–9327.

VISSER, I., RAIJMAKERS, M. E. and MOLENAAR, P. C. (2000). Confidence intervals for hidden Markov model parameters. *Br. J. Math. Stat. Psychol.* **53** 317–327.

YU, S.-Z. (2015). *Hidden Semi-Markov Models: Theory, Algorithms and Applications*. Morgan Kaufmann, San Mateo.

ZHAI, S., JACOB, D. J., PENDERGRASS, D. C., COLOMBI, N. K., SHAH, V., YANG, L. H., ZHANG, Q., WANG, S., KIM, H. et al. (2023). Coarse particulate matter air quality in East Asia: Implications for fine particulate nitrate. *Atmos. Chem. Phys.* **23** 4271–4281.

ZHANG, B., JIAO, L., XU, G., ZHAO, S., TANG, X., ZHOU, Y. and GONG, C. (2018). Influences of wind and precipitation on different-sized particulate matter concentrations (PM 2.5, PM 10, PM 2.5–10). *Meteorol. Atmos. Phys.* **130** 383–392.

ZHU, G., WEN, Y., CAO, K., HE, S. and WANG, T. (2024). A review of common statistical methods for dealing with multiple pollutant mixtures and multiple exposures. *Front. Public Health* **12** 1377685.

ZUCCHINI, W., MACDONALD, I. L. and LANGROCK, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R*, 2nd ed. *Monographs on Statistics and Applied Probability* **150**. CRC Press, Boca Raton, FL. MR3618333

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 238–259
https://doi.org/10.1214/26-AOAS2147

MANDERA: MALICIOUS NODE DETECTION IN FEDERATED LEARNING VIA RANKING

BY WANCHUANG ZHU[1,a], BENJAMIN ZI HAO ZHAO[2,b], SIMON LUO[3,c] AND KE DENG[4,d]

[1]*Centre in Data Analytics for Resources and Environment, University of Sydney,* [a]*wanchuang.zhu@sydney.edu.au*
[2]*School of Computing, Macquarie University,* [b]*ben_zi.zhao@mq.edu.au*
[3]*School of Computer Science and Engineering, The University of New South Wales,* [c]*simon.luo@unsw.edu.au*
[4]*Department of Statistics and Data Science, Tsinghua University,* [d]*kdeng@tsinghua.edu.cn*

While federated learning is a popular framework for distributed learning in the machine learning community that allows a global model to be trained across decentralized devices without data exchanging, it is vulnerable to Byzantine attacks where some involved devices are manipulated to poison the model training. Defending federated learning from various Byzantine attacks has been an active research topic in machine learning in recent years. This paper proposes a novel defense strategy called MANDERA, which achieves effective defense via precise detection of the manipulated devices based on a statistical analysis of a ranking matrix obtained from the messages reported by decentralized devices. Compared to existing defense strategies, MANDERA enjoys a higher defense efficiency against a wide range of Byzantine attacks and a clear theoretical guarantee. The effectiveness and robustness of MANDERA are further confirmed by a collection of real data analyses.

REFERENCES

BAGDASARYAN, E., VEIT, A., HUA, Y., ESTRIN, D. and SHMATIKOV, V. (2020). How to backdoor federated learning. In *International Conference on Artificial Intelligence and Statistics* 2938–2948. PMLR.

BANERJEE, M., DUROT, C. and SEN, B. (2019). Divide and conquer in nonstandard problems and the super-efficiency phenomenon. *Ann. Statist.* **47** 720–757. MR3909948 https://doi.org/10.1214/17-AOS1633

BARUCH, G., BARUCH, M. and GOLDBERG, Y. (2019). A little is enough: Circumventing defenses for distributed learning. *Adv. Neural Inf. Process. Syst.* **32**.

BLANCHARD, P., EL MHAMDI, E. M., GUERRAOUI, R. and STAINER, J. (2017). Machine learning with adversaries: Byzantine tolerant gradient descent. In *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) **30**. Curran Associates, Red Hook.

CAO, X., FANG, M., LIU, J. and GONG, N. (2021). FLTrust: Byzantine-robust federated learning via trust bootstrapping. In *Proceedings of NDSS*.

CAO, X., JIA, J. and GONG, N. Z. (2021). Provably secure federated learning against malicious clients. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

CHEN, X., JING, W., LIU, W. and ZHANG, Y. (2024). Distributed estimation and inference for semiparametric binary response models. *Ann. Statist.* **52** 922–947. MR4784064 https://doi.org/10.1214/24-aos2376

CHEN, X., LIU, W. and ZHANG, Y. (2022). First-order Newton-type estimator for distributed estimation and inference. *J. Amer. Statist. Assoc.* **117** 1858–1874. MR4528476 https://doi.org/10.1080/01621459.2021.1891925

CHEN, X. and XIE, M. (2014). A split-and-conquer approach for analysis of extraordinarily large data. *Statist. Sinica* **24** 1655–1684. MR3308656

CHEN, Y., SU, L. and XU, J. (2017). Distributed statistical machine learning in adversarial settings: Byzantine gradient descent. In *Proceedings of the ACM on Measurement and Analysis of Computing Systems* **1**.

CHEN, Z., TIAN, P., LIAO, W. and YU, W. (2021). Zero knowledge clustering based adversarial mitigation in heterogeneous federated learning. *IEEE Trans. Netw. Sci. Eng.* **8** 1070–1083. MR4291097 https://doi.org/10.1109/tnse.2020.3002796

COTTER, A., SHAMIR, O., SREBRO, N. and SRIDHARAN, K. (2011). Better mini-batch algorithms via accelerated gradient methods. *Adv. Neural Inf. Process. Syst.* **24**.

Key words and phrases. Federated learning, Byzantine attack, malicious detection, ranking.

DENG, L. (2012). The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Process. Mag.* **29** 141–142.

DOBRIBAN, E. and SHENG, Y. (2020). WONDER: Weighted one-shot distributed ridge regression in high dimensions. *J. Mach. Learn. Res.* **21** Paper No. 66, 52. MR4095345

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

FANG, M., CAO, X., JIA, J. and GONG, N. (2020). Local model poisoning attacks to Byzantine-robust federated learning. In 29*th USENIX Security Symposium* (*USENIX Security* 20) 1605–1622. USENIX Association.

FRALEY, C. and RAFTERY, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Statist. Assoc.* **97** 611–631. MR1951635 https://doi.org/10.1198/016214502760047131

GUAN, H., YAP, P., BOZOKI, A. and LIU, M. (2024). Federated learning for medical image analysis: A survey. *Pattern Recognit.* **151** 110424.

GUERRAOUI, R., ROUAULT, S. et al. (2018). The hidden vulnerability of distributed learning in Byzantium. In *International Conference on Machine Learning* 3521–3530. PMLR.

HAN, X., KASHIF, R. and ROLAND, V. (2017). Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. Preprint. Available at arXiv:1708.07747.

KAIROUZ, P., MCMAHAN, H. B., AVENT, B., BELLET, A., BENNIS, M., BHAGOJI, A. N., BONAWITZ, K., CHARLES, Z., CORMODE, G. et al. (2021). Advances and open problems in federated learning. *Found. Trends Mach. Learn.* **14** 1–210.

KLEINER, A., TALWALKAR, A., SARKAR, P. and JORDAN, M. I. (2014). A scalable bootstrap for massive data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 795–816. MR3248677 https://doi.org/10.1111/rssb.12050

KRIZHEVSKY, A., HINTON, G. et al. (2009). Learning multiple layers of features from tiny images.

LAMPORT, L., SHOSTAK, R. and PEASE, M. (2019). The Byzantine generals problem. In *Concurrency*: *The Works of Leslie Lamport* 203–226. ACM, New York.

LEE, J. D., LIU, Q., SUN, Y. and TAYLOR, J. E. (2017). Communication-efficient sparse regression. *J. Mach. Learn. Res.* **18** Paper No. 5, 30. MR3625709

LI, S., CHENG, Y., WANG, W., LIU, Y. and CHEN, T. (2020). Learning to detect malicious clients for robust federated learning. Preprint. Available at arXiv:2002.00211.

MCMAHAN, B., MOORE, E., RAMAGE, D., HAMPSON, S. and Y ARCAS, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the* 20*th International Conference on Artificial Intelligence and Statistics* (A. Singh and J. Zhu, eds.). *Proceedings of Machine Learning Research* **54** 1273–1282. PMLR.

SO, J., GÜLER, B. and AVESTIMEHR, A. S. (2021). Byzantine-resilient secure federated learning. *IEEE J. Sel. Areas Commun.* **39** 2168–2181.

STEINHARDT, J. (2018). Robust learning: Information theory and algorithms Ph.D. thesis, Stanford Univ.

TANG, K., LIU, W. and MAO, X. (2024). Multi-consensus decentralized primal-dual fixed point algorithm for distributed learning. *Mach. Learn.* **113** 4315–4357. MR4753661 https://doi.org/10.1007/s10994-024-06537-8

TOLPEGIN, V., TRUEX, S., GURSOY, M. E. and LIU, L. (2020). Data poisoning attacks against federated learning systems. In *European Symposium on Research in Computer Security* 480–501. Springer, Berlin.

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 Digit. Med.* **5** 76.

TU, J., LIU, W. and MAO, X. (2023). Byzantine-robust distributed sparse learning for *M*-estimation. *Mach. Learn.* **112** 3773–3804. MR4637806 https://doi.org/10.1007/s10994-021-06001-x

TU, J., LIU, W. and MAO, X. (2024). Distributed estimation on semi-supervised generalized linear model. *J. Mach. Learn. Res.* **25** Paper No. [76], 41. MR4749112

VOLGUSHEV, S., CHAO, S.-K. and CHENG, G. (2019). Distributed inference for quantile regression processes. *Ann. Statist.* **47** 1634–1662. MR3911125 https://doi.org/10.1214/18-AOS1730

WU, S., HUANG, D. and WANG, H. (2023). Quasi-Newton updating for large-scale distributed learning. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 1326–1354. MR4718543 https://doi.org/10.1093/jrsssb/qkad059

WU, Z., LING, Q., CHEN, T. and GIANNAKIS, G. B. (2020a). Federated variance-reduced stochastic gradient descent with robustness to Byzantine attacks. *IEEE Trans. Signal Process.* **68** 4583–4596. MR4144923 https://doi.org/10.1109/TSP.2020.3012952

WU, Z., LING, Q., CHEN, T. and GIANNAKIS, G. B. (2020b). Byrd-SAGA—GitHub. https://github.com/MrFive5555/Byrd-SAGA.

XIE, C., KOYEJO, S. and GUPTA, I. (2019). Zeno: Distributed stochastic gradient descent with suspicion-based fault-tolerance. In *International Conference on Machine Learning* 6893–6901. PMLR.

XIE, C., KOYEJO, S. and GUPTA, I. (2020). Zeno++: Robust fully asynchronous SGD. In *International Conference on Machine Learning* 10495–10503. PMLR.

YADAV, C. and BOTTOU, L. (2019). *Cold Case*: *The Lost MNIST Digits*. *In Advances in Neural Information Processing Systems* **32**. Curran Associates, Red Hook.

YIN, D., CHEN, Y., KANNAN, R. and BARTLETT, P. (2018). Byzantine-robust distributed learning: Towards optimal statistical rates. In *International Conference on Machine Learning* 5650–5659. PMLR.

ZHANG, C., YANG, S., MAO, L. and NING, H. (2024). Anomaly detection and defense techniques in federated learning: A comprehensive review. *Artif*. *Intell*. *Rev*. **57** 150.

ZHANG, Y., DUCHI, J. and WAINWRIGHT, M. (2015). Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *J*. *Mach*. *Learn*. *Res*. **16** 3299–3340. MR3450540

ZHAO, T., CHENG, G. and LIU, H. (2016). A partially linear framework for massive heterogeneous data. *Ann*. *Statist*. **44** 1400–1437. MR3519928 https://doi.org/10.1214/15-AOS1410

ZHU, W., ZHAO, B. Z, LUO, S. and DENG, K. (2026). Supplement to "MANDERA: Malicious node detection in federated learning via ranking." https://doi.org/10.1214/26-AOAS2147SUPPA, https://doi.org/10.1214/26-AOAS2147SUPPB

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 260–284
https://doi.org/10.1214/25-AOAS2024

DATA HARMONIZATION VIA REGULARIZED NONPARAMETRIC MIXING DISTRIBUTION ESTIMATION

BY STEVEN WILKINS-REEVES[1,a], YEN-CHI CHEN[1,b] AND KWUN CHUEN GARY CHAN[2,c]

[1] *Department of Statistics, University of Washington,* [a]*stevewr@uw.edu,* [b]*yenchic@uw.edu*
[2] *Department of Biostatistics, University of Washington,* [c]*kcgchan@uw.edu*

Data harmonization is the process of developing an equivalence between two measurements of a common domain. Our problem is motivated by dementia research in which multiple neuropsychological tests have been used in practice to measure the same underlying cognitive ability, such as memory or attention. We connect this statistical problem to mixing distribution estimation common in empirical Bayes approaches. We introduce and study a nonparametric latent trait model, develop a method that enforces the uniqueness of the regularized maximum likelihood estimator, show how a nonparametric EM algorithm will converge weakly to its maximizer, and illustrate its superior computational efficiency to off-the-shelf solvers. Furthermore, we develop methods for model selection and assessing the goodness-of-fit for the measurement model, an area neglected in most mixing distribution estimation problems. We develop methods for score conversion with uncertainty quantification in order to draw inferences on a whole population with multiple score scales. We apply our method to the National Alzheimer's Coordination Center Uniform Dataset and show that we can use our method to convert between score measurements and account for the measurement error. We show that this method outperforms standard techniques commonly used in dementia research.

REFERENCES

ALFONSI, A. and JOURDAIN, B. (2014). A remark on the optimal transport between two probability measures sharing the same copula. *Statist. Probab. Lett.* **84** 131–134. MR3131266 https://doi.org/10.1016/j.spl.2013.09.035

ANDERSEN, E. D. and ANDERSEN, K. D. (2000). The Mosek interior point optimizer for linear programming: An implementation of the homogeneous algorithm. In *High Performance Optimization. Appl. Optim.* **33** 197–232. Kluwer Academic, Dordrecht. MR1748773 https://doi.org/10.1007/978-1-4757-3216-0_8

BASULTO-ELIAS, G., CARRIQUIRY, A. L., DE BRABANTER, K. and NORDMAN, D. J. (2021). Bivariate kernel deconvolution with panel data. *Sankhya B* **83** 122–151. MR4256313 https://doi.org/10.1007/s13571-020-00226-x

BELL, G., LECHOWICZ, M. J. and WATERWAY, M. J. (2000). Environmental heterogeneity and species diversity of forest sedges. *J. Ecol.* **88** 67–87.

BESSER, L., KUKULL, W., KNOPMAN, D. S., CHUI, H., GALASKO, D., WEINTRAUB, S., JICHA, G., CARLSSON, C., BURNS, J. et al. (2018). Version 3 of the national Alzheimer's coordinating center's uniform data set.

CHENG, S.-T. (2016). Cognitive reserve and the prevention of dementia: The role of physical and cognitive activities. *Curr. Psychiatry Rep.* **18** 1–12.

CHERNOZHUKOV, V. and HANSEN, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *J. Econometrics* **132** 491–525. MR2323990 https://doi.org/10.1016/j.jeconom.2005.02.009

CHUNG, Y. and LINDSAY, B. G. (2015). Convergence of the EM algorithm for continuous mixing distributions. *Statist. Probab. Lett.* **96** 190–195. MR3281765 https://doi.org/10.1016/j.spl.2014.09.021

COVER, T. M. (1984). An algorithm for maximizing expected log investment return. *IEEE Trans. Inf. Theory* **30** 369–373. MR0754868 https://doi.org/10.1109/TIT.1984.1056869

Key words and phrases. Alzheimer's disease, nonparametric, expectation-maximization algorithm, latent trait model, measurement error model.

EFRON, B. (2019). Bayes, oracle Bayes and empirical Bayes. *Statist. Sci.* **34** 177–201. MR3983318 https://doi.org/10.1214/18-STS674

GRIFFITH, L., VAN DEN HEUVEL, E., FORTIER, I., HOFER, S., RAINA, P., SOHEL, N., PAYETTE, H., WOLFSON, C. and BELLEVILLE, S. (2013). Harmonization of cognitive measures in individual participant data and aggregate data meta-analysis. In *Harmonization of Cognitive Measures in Individual Participant Data and Aggregate Data Meta-Analysis* Agency for Healthcare Research and Quality (US), Rockville (MD).

GRUHL, J., EROSHEVA, E. A. and CRANE, P. K. (2013). A semiparametric approach to mixed outcome latent variable models: Estimating the association between cognition and regional brain volumes. *Ann. Appl. Stat.* **7** 2361–2383. MR3161726 https://doi.org/10.1214/13-AOAS675

GUO, F. R. and RICHARDSON, T. S. (2021). Chernoff-type concentration of empirical probabilities in relative entropy. *IEEE Trans. Inf. Theory* **67** 549–558. MR4231971 https://doi.org/10.1109/TIT.2020.3034539

IGNATIADIS, N. and WAGER, S. (2019). Covariate-powered empirical Bayes estimation. *Adv. Neural Inf. Process. Syst.* **32**.

IGNATIADIS, N. and WAGER, S. (2022). Confidence intervals for nonparametric empirical Bayes analysis. *J. Amer. Statist. Assoc.* **117** 1149–1166. MR4480697 https://doi.org/10.1080/01621459.2021.2008403

JOHNSON, M. S. (2007). Modeling dichotomous item responses with free-knot splines. *Comput. Statist. Data Anal.* **51** 4178–4192. MR2364438 https://doi.org/10.1016/j.csda.2006.04.021

JÖRESKOG, K. G. and MOUSTAKI, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivar. Behav. Res.* **36** 347–387.

KANG, H., KREUELS, B., ADJEI, O., KRUMKAMP, R., MAY, J. and SMALL, D. S. (2013). The causal effect of malaria on stunting: A Mendelian randomization and matching approach. *Int. J. Epidemiol.* **42** 1390–1398.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906. MR0086464 https://doi.org/10.1214/aoms/1177728066

KOSMOL, P. and MÜLLER-WICHARDS, D. (2011). *Optimization in Function Spaces*: *With Stability Considerations in Orlicz Spaces*. *De Gruyter Series in Nonlinear Analysis and Applications* **13**. de Gruyter, Berlin. MR2760903

LAIRD, N. (1978). Nonparametric maximum likelihood estimation of a mixed distribution. *J. Amer. Statist. Assoc.* **73** 805–811. MR0521328

LINDSAY, B., CLOGG, C. C. and GREGO, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *J. Amer. Statist. Assoc.* **86** 96–107. MR1137102

LINDSAY, B. G. (1983a). The geometry of mixture likelihoods. II. The exponential family. *Ann. Statist.* **11** 783–792. MR0707929 https://doi.org/10.1214/aos/1176346245

LINDSAY, B. G. (1983b). The geometry of mixture likelihoods: A general theory. *Ann. Statist.* **11** 86–94. MR0684866 https://doi.org/10.1214/aos/1176346059

LINDSAY, B. G. (1995). Mixture models: Theory, geometry and applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics* i–163. JSTOR.

LORD, F. M. (1965). A strong true-score theory, with applications. *Psychometrika* **30** 239–270.

LORD, F. M. (1969). Estimating true-score distributions in psychological testing (an empirical Bayes estimation problem). *Psychometrika* **34** 259–299.

MARDIA, J., JIAO, J., TÁNCZOS, E., NOWAK, R. D. and WEISSMAN, T. (2018). Concentration inequalities for the empirical distribution.

MENG, X. and D'ARCY, C. (2012). Education and dementia in the context of the cognitive reserve hypothesis: A systematic review with meta-analyses and qualitative analyses. *PLoS ONE* **7** e38268.

MEREDITH, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* **58** 525–543. MR1248361 https://doi.org/10.1007/BF02294825

MONSELL, S. E., DODGE, H. H., ZHOU, X. H., BU, Y., BESSER, L. M., MOCK, C., HAWES, S. E., KUKULL, W. A., WEINTRAUB, S. et al. (2016). Results from the NACC uniform data set neuropsychological battery crosswalk study. *Alzheimer Dis. Assoc. Disord.* **30** 134–139.

O'DONOGHUE, B., CHU, E., PARIKH, N. and BOYD, S. (2016). Conic optimization via operator splitting and homogeneous self-dual embedding. *J. Optim. Theory Appl.* **169** 1042–1068. MR3501397 https://doi.org/10.1007/s10957-016-0892-3

PAGANIN, S., PACIOREK, C. J., WEHRHAHN, C., RODRIGUEZ, A., RABE-HESKETH, S. and DE VALPINE, P. (2021). Computational methods for Bayesian semiparametric item response theory models.

RASCH, G. (1966). An individualistic approach to item analysis Technical Report Readings in Mathematical Social Science.

ROBINS, J. M. and TSIATIS, A. A. (1991). Correcting for noncompliance in randomized trials using rank preserving structural failure time models. *Comm. Statist. Theory Methods* **20** 2609–2631. MR1144866 https://doi.org/10.1080/03610929108830654

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196 https://doi.org/10.1093/biomet/63.3.581

RUBIN, D. B. (1996). Multiple imputation after 18+ years. *J. Amer. Statist. Assoc.* **91** 473–489.

SCHOMAKER, M. and HEUMANN, C. (2018). Bootstrap inference when using multiple imputation. *Stat. Med.* **37** 2252–2266. MR3810720 https://doi.org/10.1002/sim.7654

SIENSKI, G., NARAYAN, P., BONNER, J. M., KORY, N., BOLAND, S., ARCZEWSKA, A. A., RALVENIUS, W. T., AKAY, L., LOCKSHIN, E. et al. (2021). APOE4 disrupts intracellular lipid homeostasis in human iPSC-derived glia. *Sci. Transl. Med.* **13** eaaz4564.

SILVERMAN, B. W., JONES, M. C., NYCHKA, D. W. and WILSON, J. D. (1990). A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography. *J. Roy. Statist. Soc. Ser. B, Methodol.* **52** 271–324. MR1064419

STEFANSKI, L. and CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184. MR1054861 https://doi.org/10.1080/02331889008802238

STERN, Y. (2009). Cognitive reserve. *Neuropsychologia* **47** 2015–2028.

STERN, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol.* **11** 1006–1012.

TIAN, K., KONG, W. and VALIANT, G. (2017). Learning populations of parameters. In *Advances in Neural Information Processing Systems* **30**.

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

VAN DEN HEUVEL, E. R., GRIFFITH, L. E., SOHEL, N., FORTIER, I., MUNIZ-TERRERA, G. and RAINA, P. (2020). Latent variable models for harmonization of test scores: A case study of memory. *Biom. J.* **62** 34–52. MR4052772 https://doi.org/10.1002/bimj.201800146

VARDI, Y., SHEPP, L. A. and KAUFMAN, L. (1985). A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.* **80** 8–37. MR0786595

VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 https://doi.org/10.1090/gsm/058

VINAYAK, R. K., KONG, W., VALIANT, G. and KAKADE, S. M. (2019). Maximum Likelihood Estimation for Learning Populations of Parameters. In: *36th International Conference on Machine Learning, ICML 2019*, 2019-June, 11217–11226.

WANG, L., ROBINS, J. M. and RICHARDSON, T. S. (2017). On falsification of the binary instrumental variable model. *Biometrika* **104** 229–236. MR3626478 https://doi.org/10.1093/biomet/asx011

WEINTRAUB, S., SALMON, D., MERCALDO, N., FERRIS, S., GRAFF-RADFORD, N. R., CHUI, H., CUMMINGS, J., DECARLI, C., FOSTER, N. L. et al. (2009). The Alzheimer's disease centers' uniform data set (UDS): The neuropsychologic test battery. *Alzheimer Dis. Assoc. Disord.* **23** 91–101.

WHITE, I. R., BABIKER, A. G., WALKER, S. and DARBYSHIRE, J. H. (1999). Randomization-based methods for correcting for treatment changes: Examples from the Concorde trial. *Stat. Med.* **18** 2617–2634.

WHITE, I. R., WALKER, S., BABIKER, A. G. and DARBYSHIRE, J. H. (1997). Impact of treatment changes on the interpretation of the Concorde trial. *AIDS* **11** 999–1006.

WILKINS-REEVES, S., CHEN, Y.-C and CHAN, K. C (2026). Supplement to "Data Harmonization via Regularized nonparametric Mixing Distribution Estimation." https://doi.org/10.1214/25-AOAS2024SUPPA, https://doi.org/10.1214/25-AOAS2024SUPPB

WOOD, G. R. (1999). Binomial mixtures: Geometric estimation of the mixing distribution. *Ann. Statist.* **27** 1706–1721. MR1742506 https://doi.org/10.1214/aos/1017939148

WOODS, C. M. and THISSEN, D. (2006). Item response theory with estimation of the latent population distribution using spline-based densities. *Psychometrika* **71** 281–301. MR2259176 https://doi.org/10.1007/s11336-004-1175-8

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 285–306
https://doi.org/10.1214/25-AOAS2078
© Institute of Mathematical Statistics, 2026

B-BIND: BIOPHYSICAL BAYESIAN INFERENCE FOR NEURODEGENERATIVE DYNAMICS

BY ANAMIKA AGRAWAL[1,6,a], VICTORIA M. RACHLEFF[2,4,b], KYLE J. TRAVAGLINI[2,c], SHUBHABRATA MUKHERJEE[5,i], PAUL K. CRANE[5,j], MICHAEL HAWRYLYCZ[2,d], C. DIRK KEENE[4,h], ED LEIN[2,e], GONZALO E. MENA[3,g] AND MARIANO I. GABITTO[2,7,f]

[1]*Center for Data-Driven Discovery for Biology, Allen Institute,* [a]*anamika.agrawal@alleninstitute.org*

[2]*Human Cell Types Department, Allen Institute,* [b]*victoria.rachleff@alleninstitute.org,* [c]*kyle.travaglini@alleninstitute.org,* [d]*mikeh@alleninstitute.org,* [e]*edl@alleninstitute.org,* [f]*mariano.gabitto@alleninstitute.org*

[3]*Department of Statistics & Data Science, Carnegie Mellon University,* [g]*gmena@andrew.cmu.edu*

[4]*Department of Laboratory Medicine and Pathology, University of Washington,* [h]*cdkeene@uw.edu*

[5]*Department of Medicine, University of Washington,* [i]*smukherj@uw.edu,* [j]*pcrane@uw.edu*

[6]*Department of Neurobiology and Biophysics, University of Washington*

[7]*Department of Statistics, University of Washington*

Throughout an organism's life, numerous complex and interdependent biological systems undergo transitions driven by biophysical processes. These processes reflect the underlying biological state and serve as measurable indicators of the organism's condition. A central objective in modern biology and neuroscience is to infer these latent, unobserved states and to reconstruct the trajectories these systems follow over time. However, in many experimental settings, we are limited to discrete snapshots—observations captured at different time points across different individuals—which complicates the task of recovering the continuous underlying trajectory. This challenge is particularly relevant in the study of Alzheimer's disease (AD) progression, where we can measure the aggregation of pathological proteins in postmortem brain samples, but the true course of disease remains hidden.

This paper proposes a biophysically motivated Bayesian framework (B-BIND: Biophysical Bayesian Inference for Neurodegenerative Dynamics), where the disease state is modeled and inferred from observed AD pathological proteins. Inspired by biophysical models, we describe pathological burden as an exponential process. The progression of AD is modeled by a latent variable, termed pseudotime, creating a pseudotemporal order of donors based on their pathological burden. We study the theoretical properties of the model using linearization to reveal convergence and identifiability properties. We provide Markov chain Monte Carlo estimation algorithms, illustrating the effectiveness of our approach with multiple simulation studies across various data conditions. Applying this methodology to data from the Seattle Alzheimer's Disease Brain Cell Atlas, we infer pseudotime of donors to then refine the model, focusing on the most informative pathologies. This framework lays the groundwork for continuous pseudotime modeling in the analysis of neurodegenerative diseases.

REFERENCES

AGRAWAL, A., RACHLEFF, V. M., TRAVAGLINI, K. J., MUKHERJEE, S., CRANE, P. K., HAWRYLYCZ, M., KEENE, C. D., LEIN, E., MENA, G. E. and GABITTO, M. I. (2026). Supplement to "B-BIND: Biophysical Bayesian Inference for Neurodegenerative Dynamics." https://doi.org/10.1214/25-AOAS2078SUPP

BAI, J. and NG, S. (2013). Principal components estimation and identification of static factors. *J. Econometrics* **176** 18–29. MR3067022 https://doi.org/10.1016/j.jeconom.2013.03.007

Key words and phrases. Hierarchical Bayesian inference, noisy matrix factorization, identifiable, Alzheimer's disease, trajectory inference.

BERAHA, M., METELLI, A. M., PAPINI, M., TIRINZONI, A. and RESTELLI, M. (2019). Feature selection via mutual information: New theoretical insights. In 2019 *International Joint Conference on Neural Networks* (*IJCNN*) 1–9.

BINGHAM, E., CHEN, J. P., JANKOWIAK, M., OBERMEYER, F., PRADHAN, N., KARALETSOS, T., SINGH, R., SZERLIP, P. A., HORSFALL, P. et al. (2019). Pyro: Deep universal probabilistic programming. *J. Mach. Learn. Res.* **20** 28:1–28:6.

BRAAK, H., ALAFUZOFF, I., ARZBERGER, T., KRETZSCHMAR, H. and DEL TREDICI, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol.* **112** 389–404.

BRAAK, H. and BRAAK, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82** 239–259.

CAI, T. T. and ZHANG, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Ann. Statist.* **46** 60–89. MR3766946 https://doi.org/10.1214/17-AOS1541

CAMPBELL, K. R. and YAU, C. (2016). Order under uncertainty: Robust differential expression analysis using probabilistic models for pseudotime inference. *PLoS Comput. Biol.* **12** e1005212.

CAMPBELL, K. R. and YAU, C. (2018). Uncovering pseudotemporal trajectories with covariates from single cell and bulk expression data. *Nat. Commun.* **9** 2442.

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

CHEN, Y., LI, X. and ZHANG, S. (2020). Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *J. Amer. Statist. Assoc.* **115** 1756–1770. MR4189755 https://doi.org/10.1080/01621459.2019.1635485

CRANE, P. K., GIBBONS, L. E., JOLLEY, L. and VAN BELLE, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Med. Care* **44** S115–S123.

DAVIS, J. K. and SINDI, S. S. (2016). A mathematical model of the dynamics of prion aggregates with chaperone-mediated fragmentation. *J. Math. Biol.* **72** 1555–1578. MR3483184 https://doi.org/10.1007/s00285-015-0921-0

DU, J.-H., WASSERMAN, L. and ROEDER, K. (2023). Simultaneous inference for generalized linear models with unmeasured confounders. arXiv preprint. Available at arXiv:2309.07261.

DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222. MR3960671 https://doi.org/10.1016/0370-2693(87)91197-x

GABITTO, M. I., TRAVAGLINI, K. J., RACHLEFF, V. M., KAPLAN, E. S., LONG, B., ARIZA, J., DING, Y., MAHONEY, J. T., DEE, N. et al. (2023). Integrated multimodal cell atlas of Alzheimer's disease. *Res. Sq.*.

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2014). *Bayesian Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3235677

GELMAN, A. and HILL, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge Univ. Press, Cambridge.

GELMAN, A., HILL, J. and VEHTARI, A. (2021). *Regression and Other Stories*. Cambridge Univ. Press, Cambridge.

GUPTA, A. and BAR-JOSEPH, Z. (2008). Extracting dynamics from static cancer expression data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **5** 172–182.

HOEGH, A. and ROBERTS, D. W. (2020). Evaluating and presenting uncertainty in model-based unconstrained ordination. *Ecol. Evol.* **10** 59–69.

HOFFMAN, M. D. and GELMAN, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623. MR3214779

HOU, W., JI, Z., CHEN, Z., WHERRY, E. J., HICKS, S. C. and JI, H. (2023). A statistical framework for differential pseudotime analysis with multiple single-cell RNA-seq samples. *Nat. Commun.* **14** 7286.

HUANG, K., ZHANG, Y., GONG, H., QIAO, Z., WANG, T., ZHAO, W., HUANG, L. and ZHOU, X. (2023). Inferring evolutionary trajectories from cross-sectional transcriptomic data to mirror lung adenocarcinoma progression. *PLoS Comput. Biol.* **19** e1011122.

HUI, F. K., WARTON, D. I., FOSTER, S. D. and HAAK, C. R. (2023). Spatiotemporal joint species distribution modelling: A basis function approach. *Methods Ecol. Evol.* **14** 2150–2164.

HUI, F. K. C., TASKINEN, S., PLEDGER, S., FOSTER, S. D. and WARTON, D. I. (2015). Model-based approaches to unconstrained ordination. *Methods Ecol. Evol.* **6** 399–411.

HYMAN, B., PHELPS, C., BEACH, T., BIGIO, E., CAIRNS, N., CARRILLO, M., DICKSON, D., DUYCKAERTS, C., FROSCH, M. et al. (2012). National institute on aging-Alzheimer's association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's Dement.* **8**.

KIDZIŃSKI, Ł., HUI, F. K. C., WARTON, D. I. and HASTIE, T. J. (2022). Generalized matrix factorization: Efficient algorithms for fitting generalized linear latent variable models to large data arrays. *J. Mach. Learn. Res.* **23** Paper No. [291]. MR4577730

KIM, S. and CAMILLI, G. (2014). An item response theory approach to longitudinal analysis with application to summer setback in preschool language/literacy. *Large-Scale Assess. Educ.* **2** 1–17.

LIU, W., LIN, H., ZHENG, S. and LIU, J. (2023). Generalized factor model for ultra-high dimensional correlated variables with mixed types. *J. Amer. Statist. Assoc.* **118** 1385–1401. MR4595502 https://doi.org/10.1080/01621459.2021.1999818

MAGWENE, P. M., LIZARDI, P. and KIM, J. (2003). Reconstructing the temporal ordering of biological samples using microarray data. *Bioinformatics* **19** 842–850.

MASEL, J., JANSEN, V. A. and NOWAK, M. A. (1999). Quantifying the kinetic parameters of prion replication. *Biophys. Chem.* **77** 139–152.

MASTERS, C. L., BATEMAN, R., BLENNOW, K., ROWE, C. C., SPERLING, R. A. and CUMMINGS, J. L. (2015). Alzheimer's disease. *Nat. Rev. Dis. Primers* **1** 1–18.

MUKHERJEE, S., HEATH, L., PREUSS, C., JAYADEV, S., GARDEN, G. A., GREENWOOD, A. K., SIEBERTS, S. K., DE JAGER, P. L., ERTEKIN-TANER, N. et al. (2020). Molecular estimation of neurodegeneration pseudotime in older brains. *Nat. Commun.* **11** 5781.

NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

O'HARA, R. B. and VAN DER VEEN, B. (2024). Hierarchical ordination, a unifying framework for drivers of community processes. *bioRxiv* 2024–01.

PHAN, D., PRADHAN, N. and JANKOWIAK, M. (2019). Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. arXiv preprint. Available at arXiv:1912.11554.

PIERSON, E., KOH, P., HASHIMOTO, T., KOLLER, D., LESKOVEC, J., ERIKSSON, N. and LIANG, P. (2019). Inferring multidimensional rates of aging from cross-sectional data. *Proc. Mach. Learn. Res.* **89**.

POPOVIC, G. C., HUI, F. K. and WARTON, D. I. (2022). Fast model-based ordination with copulas. *Methods Ecol. Evol.* **13** 194–202.

REID, J. E. and WERNISCH, L. (2016). Pseudotime estimation: Deconfounding single cell time series. *Bioinformatics* **32** 2973–2980.

ROBERTS, D. W. (2020). Comparison of distance-based and model-based ordinations. *Ecology* **101** e02908.

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

STREET, K., RISSO, D., FLETCHER, R. B., DAS, D., NGAI, J., YOSEF, N., PURDOM, E. and DUDOIT, S. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19** 1–16.

TENG, E., HASEGAWA, K., HOMMA, A., IMAI, Y., LARSON, E., GRAVES, A., SUGIMOTO, K., YAMAGUCHI, T., SASAKI, H. et al. (1994). The cognitive abilities screening instrument (CASI): A practical test for cross-cultural epidemiological studies of dementia. *Int. Psychogeriatr.* **6**.

THAL, D. R., RÜB, U., ORANTES, M. and BRAAK, H. (2002). Phases of Aβ-deposition in the human brain and its relevance for the development of AD. *Neurology* **58** 1791–1800.

TIJMS, B. M., VROMEN, E. M., MJAAVATTEN, O. et al. (2024). Cerebrospinal fluid proteomics in patients with Alzheimer's disease reveals five molecular subtypes with distinct genetic risk profiles. *Nat. Aging* **4** 33–47.

TOMPA, P., TUSNÁDY, G., CSERZŐ, M. and SIMON, I. (2001). Prion protein: Evolution caught en route. *Proc. Natl. Acad. Sci. USA* **98** 4431–4436.

TRAPNELL, C., CACCHIARELLI, D., GRIMSBY, J., POKHAREL, P., LI, S., MORSE, M., LENNON, N. J., LIVAK, K. J., MIKKELSEN, T. S. et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32** 381–386.

VAN DER VEEN, B., HUI, F. K., HOVSTAD, K. A. and O'HARA, R. B. (2023). Concurrent ordination: Simultaneous unconstrained and constrained latent variable modelling. *Methods Ecol. Evol.* **14** 683–695.

VAQUER-ALICEA, J. and DIAMOND, M. I. (2019). Propagation of protein aggregation in neurodegenerative diseases. *Annu. Rev. Biochem.* **88** 785–810.

VEHTARI, A., GELMAN, A. and GABRY, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27** 1413–1432. MR3647105 https://doi.org/10.1007/s11222-016-9696-4

VOGEL, J. and HANSOON, O. (2022). Subtypes of Alzheimer's disease: Questions, controversy, and meaning. *Trends Neurosci.* **45** 342–345.

WADDINGTON, C. H. (1940). *Organisers and Genes*. Cambridge Univ. Press, Cambridge.

WADDINGTON, C. H. (1957). *The Strategy of Genes*. Taylor & Francis, London.

WANG, F. (2022). Maximum likelihood estimation and inference for high dimensional generalized factor models with application to factor-augmented regressions. *J. Econometrics* **229** 180–200. MR4414018 https://doi.org/10.1016/j.jeconom.2020.11.002

WATANABE, S. and OPPER, M. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11** 3571–3594. MR2756194

WEDIN, P. (1972). Perturbation bounds in connection with singular value decomposition. *BIT* **12** 99–111. MR0309968 https://doi.org/10.1007/bf01932678

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 307–321
https://doi.org/10.1214/25-AOAS2086

A BAYESIAN JOINT MODEL OF MULTIPLE LONGITUDINAL AND CATEGORICAL OUTCOMES WITH APPLICATION TO MULTIPLE MYELOMA USING PERMUTATION-BASED VARIABLE IMPORTANCE

BY DANILO ALVARES[1,a], JESSICA K. BARRETT[1,b], FRANÇOIS MERCIER[2,c], JOCHEN SCHULZE[2,d], SEAN YIU[3,h], FELIPE CASTRO[2,e], SPYROS ROUMPANIS[2,f] AND YAJING ZHU[2,g]

[1]*MRC Biostatistics Unit, University of Cambridge*, [a]*danilo.alvares@mrc-bsu.cam.ac.uk*, [b]*jessica.barrett@mrc-bsu.cam.ac.uk*
[2]*F. Hoffmann-La Roche Ltd*, [c]*francois.mercier@roche.com*, [d]*schulzejochen797@gmail.com*, [e]*felipe.castro@roche.com*, [f]*spyros.roumpanis@roche.com*, [g]*yajing.zhu09@gmail.com*
[3]*Roche Products Ltd*, [h]*sean_yiu@hotmail.com*

Joint models have proven to be an effective approach for uncovering potentially hidden connections between various types of outcomes, mainly continuous, time-to-event, and binary. Typically, longitudinal continuous outcomes are characterized by linear mixed-effects models, survival outcomes are described by proportional hazards models, and the link between outcomes are captured by shared random effects. Other modeling variations include generalized linear mixed-effects models for longitudinal data and logistic regression when a binary outcome is present, rather than time until an event of interest. However, in a clinical research setting, one might be interested in modeling the physician's chosen treatment based on the patient's medical history to identify prognostic factors. In this situation there are often multiple treatment options, requiring the use of a multiclass classification approach. Inspired by this context, we develop a Bayesian joint model for longitudinal and categorical data. In particular, our motivation comes from a multiple myeloma study in which biomarkers display nonlinear trajectories that are well captured through biexponential submodels, where patient-level information is shared with the categorical submodel. We also present a variable importance strategy to rank prognostic factors. We apply our proposal and a competing model to the multiple myeloma data, compare the variable importance and inferential results for both models, and illustrate patient-level interpretations using our joint model.

REFERENCES

AGRESTI, A. (2013). *Categorical Data Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3087436

ALSEFRI, M., SUDELL, M., GARCÍA-FIÑANA, M. and KOLAMUNNAGE-DONA, R. (2020). Bayesian joint modelling of longitudinal and time to event data: A methodological review. *BMC Med. Res. Methodol.* **20** 1–17.

ALVARES, D., BARRETT, J. K., MERCIER, F., ROUMPANIS, S., YIU, S., CASTRO, F., SCHULZE, J. and ZHU, Y. (2025). A Bayesian joint model of multiple nonlinear longitudinal and competing risks outcomes for dynamic prediction in multiple myeloma: Joint estimation and corrected two-stage approaches. *Stat. Med.* **44** Paper No. e10322. MR4860475 https://doi.org/10.1002/sim.10322

ALVARES, D., BARRETT, J. K., MERCIER, F., SCHULZE, J., YIU, S., CASTRO, F., ROUMPANIS, S. and ZHU, Y. (2026). Supplement to "A Bayesian joint model of multiple longitudinal and categorical outcomes with application to multiple myeloma using permutation-based variable importance." https://doi.org/10.1214/25-AOAS2086SUPPA, https://doi.org/10.1214/25-AOAS2086SUPPB

ALVARES, D. and RUBIO, F. J. (2021). A tractable Bayesian joint model for longitudinal and survival data. *Stat. Med.* **40** 4213–4229. MR4300082 https://doi.org/10.1002/sim.9024

Key words and phrases. Bayesian inference, biexponential model, categorical data, model-agnostic, M-protein dynamics.

ANDRINOPOULOU, E.-R., NASSERINEJAD, K., SZCZESNIAK, R. and RIZOPOULOS, D. (2020). Integrating latent classes in the Bayesian shared parameter joint model of longitudinal and survival outcomes. *Stat. Methods Med. Res.* **29** 3294–3307. MR4156855 https://doi.org/10.1177/0962280220924680

ANDRINOPOULOU, E.-R. and RIZOPOULOS, D. (2016). Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. *Stat. Med.* **35** 4813–4823. MR3554995 https://doi.org/10.1002/sim.7027

ANDRINOPOULOU, E.-R., RIZOPOULOS, D., TAKKENBERG, J. J. M. and LESAFFRE, E. (2014). Joint modeling of two longitudinal outcomes and competing risk data. *Stat. Med.* **33** 3167–3178. MR3260535 https://doi.org/10.1002/sim.6158

ANDRINOPOULOU, E.-R., RIZOPOULOS, D., TAKKENBERG, J. J. M. and LESAFFRE, E. (2017). Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Stat. Methods Med. Res.* **26** 1787–1801. MR3687178 https://doi.org/10.1177/0962280215588340

BIRNBAUM, B., NUSSBAUM, N., SEIDL-RATHKOPF, K., AGRAWAL, M., ESTEVEZ, M., ESTOLA, E., HAIMSON, J., HE, L., LARSON, P. et al. (2020). Model-assisted cohort selection with bias analysis for generating large-scale cohorts from the EHR for oncology research. Available at arXiv:2001.09765.

BRILLEMAN, S. L., CROWTHER, M. J., MORENO-BETANCUR, M., BUROS NOVIK, J., DUNYAK, J., AL-HUNITI, N., FOX, R., HAMMERBACHER, J. and WOLFE, R. (2019). Joint longitudinal and time-to-event models for multilevel hierarchical data. *Stat. Methods Med. Res.* **28** 3502–3515. MR4003604 https://doi.org/10.1177/0962280218808821

BROWN, E. R. (2009). Assessing the association between trends in a biomarker and risk of event with an application in pediatric HIV/AIDS. *Ann. Appl. Stat.* **3** 1163–1182. MR2750391 https://doi.org/10.1214/09-AOAS251

CASALICCHIO, G., MOLNAR, C. and BISCHL, B. (2019). Visualizing the feature importance for black box models. In *Machine Learning and Knowledge Discovery in Databases* 655–670. Springer, Cham, Switzerland.

CHAMMA, A., THIRION, B. and ENGEMANN, D. (2024). Variable importance in high-dimensional settings requires grouping. In *Proceedings of the AAAI Conference on Artificial Intelligence* 11195–11203.

CHEN, S., ALVARES, D., PALMA, M. and BARRETT, J. K. (2025). Bayesian shared parameter joint models for heterogeneous populations. *Stat. Comput.* **35** Paper No. 125. MR4920066 https://doi.org/10.1007/s11222-025-10647-1

CHEN, Z., XIAO, F., GUO, F. and YAN, J. (2023). Interpretable machine learning for building energy management: A state-of-the-art review. *Adv. Appl. Energy* **9** 1–19.

CHI, M., WANG, X., SONG, H., PENG, Y. and TU, D. (2025). Joint analysis of longitudinal ordinal categorical item response data and survival times with cure fraction. *Stat. Biopharm. Res.* **17** 67–77.

CHO, S., PSIODA, M. A. and IBRAHIM, J. G. (2024). Bayesian joint modeling of multivariate longitudinal and survival outcomes using Gaussian copulas. *Biostatistics* **25** 962–977. MR4808867 https://doi.org/10.1093/biostatistics/kxae009

CHOI, J., CAI, J., ZENG, D. and OLSHAN, A. F. (2015). Joint analysis of survival time and longitudinal categorical outcomes. *Stat. Biosci.* **7** 19–47.

CLARET, L., GIRARD, P., HOFF, P. M., VAN CUTSEM, E., ZUIDEVELD, K. P., JORGA, K., FAGERBERG, J. and BRUNO, R. (2009). Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J. Clin. Oncol.* **27** 4103–4108.

DELPORTE, M., MOLENBERGHS, G., FIEUWS, S. and VERBEKE, G. (2025). A joint normal-ordinal (probit) model for ordinal and continuous longitudinal data. *Biostatistics* **26** Paper No. kxae014. MR4865814 https://doi.org/10.1093/biostatistics/kxae014

DESMÉE, S., MENTRÉ, F., VEYRAT-FOLLET, C., SÉBASTIEN, B. and GUEDJ, J. (2017). Using the SAEM algorithm for mechanistic joint models characterizing the relationship between nonlinear PSA kinetics and survival in prostate cancer patients. *Biometrics* **73** 305–312. MR3632376 https://doi.org/10.1111/biom.12537

EMMANUEL, T., MAUPONG, T., MPOELENG, D., SEMONG, T., MPHAGO, B. and TABONA, O. (2021). A survey on missing data in machine learning. *J. Big Data* **8** 1–37.

FISHER, A., RUDIN, C. and DOMINICI, F. (2019). All models are wrong, but *many* are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* **20** Paper No. 177. MR4048988

GANJALI, M. and BAGHFALAKI, T. (2015). A copula approach to joint modeling of longitudinal measurements and survival times using Monte Carlo expectation-maximization with application to AIDS studies. *J. Biopharm. Statist.* **25** 1077–1099.

GRAN, C., AFRAM, G., LIWING, J., VERHOEK, A. and NAHI, H. (2021). Involved free light chain: An early independent predictor of response and progression in multiple myeloma. *Leuk. Lymphoma* **62** 2227–2234.

GRANDINI, M., BAGLI, E. and VISANI, G. (2020). Metrics for multi-class classification: an overview. Available at arXiv:2008.05756.

GULLA, A. and ANDERSON, K. C. (2020). Multiple myeloma: The (r)evolution of current therapy and a glance into future. *Haematologica* **105** 2358–2367.

GUO, S., ZHANG, J. and HALABI, S. (2024). Joint modelling of longitudinal measurements and time-to-event outcomes with a cure fraction using functional principal component analysis. *Stat. Med.* **43** 6059–6072. MR4842065 https://doi.org/10.1002/sim.10289

HICKEY, G. L., PHILIPSON, P., JORGENSEN, A. and KOLAMUNNAGE-DONA, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: Recent developments and issues. *BMC Med. Res. Methodol.* **16** 1–15.

HORROCKS, J. and VAN DEN HEUVEL, M. J. (2009). Prediction of pregnancy: A joint model for longitudinal and binary data. *Bayesian Anal.* **4** 523–538. MR2551044 https://doi.org/10.1214/09-BA419

HU, J. and SZYMCZAK, S. (2023). A review on longitudinal data analysis with random forest. *Brief. Bioinform.* **24** 1–11.

IBRAHIM, J. G., CHU, H. and CHEN, L. M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *J. Clin. Oncol.* **28** 2796–2801.

ISLAM, M., DANIELS, M. J., AGHABAZAZ, Z. and SIDDIQUE, J. (2024). Bayesian feature selection in joint models with application to a cardiovascular disease cohort study. Available at arXiv:2412.00885.

KOLO, B. (2010). *Binary and Multiclass Classification*, 1st ed. Weatherford Press, Weatherford, OK, USA.

KUMAR, S. K., RAJKUMAR, V., KYLE, R. A., VAN DUIN, M., SONNEVELD, P., MATEOS, M. V., GAY, F. and ANDERSON, K. C. (2017). Multiple myeloma. *Nat. Rev. Dis. Primers* **3** 1–20.

LU, X., HUANG, Y. and ZHOU, R. (2016). Joint analysis of nonlinear heterogeneous longitudinal data and binary outcome: An application to AIDS clinical studies. *J. Appl. Stat.* **43** 2713–2728. MR3546110 https://doi.org/10.1080/02664763.2016.1142951

MA, X., LONG, L., MOON, S., ADAMSON, B. J. S. and BAXI, S. S. (2023). Comparison of population characteristics in real-world clinical oncology databases in the US: Flatiron Health, SEER, and NPCR. MedRxiv:10.1101/2020.03.16.20037143v3.

MARTINS, R., SILVA, G. L. and ANDREOZZI, V. (2016). Bayesian joint modeling of longitudinal and spatial survival AIDS data. *Stat. Med.* **35** 3368–3384. MR3528263 https://doi.org/10.1002/sim.6937

MARTINS, R., SILVA, G. L. and ANDREOZZI, V. (2017). Joint analysis of longitudinal and survival AIDS data with a spatial fraction of long-term survivors: A Bayesian approach. *Biom. J.* **59** 1166–1183. MR3731209 https://doi.org/10.1002/bimj.201600159

NICODEMUS, K. K., MALLEY, J. D., STROBL, C. and ZIEGLER, A. (2010). The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC Bioinform.* **11** 1–13.

PROUST-LIMA, C., SÉNE, M., TAYLOR, J. M. G. and JACQMIN-GADDA, H. (2014). Joint latent class models for longitudinal and time-to-event data: A review. *Stat. Methods Med. Res.* **23** 74–90. MR3190688 https://doi.org/10.1177/0962280212445839

PUNKE, A. P., WADDELL, J. A. and SOLIMANDO, D. A. (2017). Lenalidomide, Bortezomib, and Dexamethasone (RVD) regimen for multiple myeloma. *Hosp. Pharm.* **52** 27–32.

R CORE TEAM (2023a). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available at https://www.R-project.org/.

RAJKUMAR, S. V. and KUMAR, S. (2020). Multiple myeloma current treatment algorithms. *Blood Cancer J.* **10** 1–10.

RAPPL, A., KNEIB, T., LANG, S. and BERGHERR, E. (2023). Spatial joint models through Bayesian structured piecewise additive joint modelling for longitudinal and time-to-event data. *Stat. Comput.* **33** Paper No. 135. MR4654175 https://doi.org/10.1007/s11222-023-10293-5

RIBEIRO, M. T., SINGH, S. and GUESTRIN, C. (2016). Model-agnostic interpretability of machine learning. Available at arXiv:1606.05386.

RIZOPOULOS, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* **67** 819–829. MR2829256 https://doi.org/10.1111/j.1541-0420.2010.01546.x

RIZOPOULOS, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data*: *With Applications in R*, 1st ed. CRC Press/CRC, Boca Raton, FL, USA.

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., PAPAGEORGIOU, G. and MORGAN, T. M. (2024). Using joint models for longitudinal and time-to-event data to investigate the causal effect of salvage therapy after prostatectomy. *Stat. Methods Med. Res.* **33** 894–908. MR4736118 https://doi.org/10.1177/09622802241239003

RUÉ, M., ANDRINOPOULOU, E.-R., ALVARES, D., ARMERO, C., FORTE, A. and BLANCH, L. (2017). Bayesian joint modeling of bivariate longitudinal and competing risks data: An application to study patient-ventilator asynchronies in critical care patients. *Biom. J.* **59** 1184–1203. MR3731210 https://doi.org/10.1002/bimj.201600221

Rustand, D., van Niekerk, J., Krainski, E. T., Rue, H. and Proust-Lima, C. (2024). Fast and flexible inference for joint models of multivariate longitudinal and survival data using integrated nested Laplace approximations. *Biostatistics* **25** 429–448. MR4732241 https://doi.org/10.1093/biostatistics/kxad019

Schuurman, N. K., Grasman, R. P. P. P. and Hamaker, E. L. (2016). A comparison of inverse-Wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivar. Behav. Res.* **51** 185–206.

Stan Development Team (2023b). RStan: The R interface to Stan. Stan. Available at http://mc-stan.org/.

Stein, W. D., Figg, W. D., Dahut, W., Stein, A. D., Hoshen, M. B., Price, D., Bates, S. E. and Fojo, T. (2008). Tumor growth rates derived from data for patients in a clinical trial correlate strongly with patient survival: A novel strategy for evaluation of clinical trial data. *The Oncologist* **13** 1046–1054.

Sun, J. and Basu, S. (2024). Penalized joint models of high-dimensional longitudinal biomarkers and a survival outcome. *Ann. Appl. Stat.* **18** 1490–1505. MR4728676 https://doi.org/10.1214/23-aoas1844

Tacchetti, P., Pezzi, A., Zamagni, E., Pantani, L., Rocchi, S., Zannetti, B. A., Mancuso, K., Ilaria Rizzello, I. and Cavo, M. (2017). Role of serum free light chain assay in the detection of early relapse and prediction of prognosis after relapse in multiple myeloma patients treated upfront with novel agents. *Haematologica* **102** 104–107.

Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N. and Asadpour, M. (2020). Boosting methods for multi-class imbalanced data classification: An experimental review. *J. Big Data* **7** 1–47.

Thai, H. T., Gaudel, N., Cerou, M., Ayral, G., Fau, J. B., Sebastien, B., van de Velde, H., Semiond, D. and Veyrat-Follet, C. (2022). Joint modelling and simulation of M-protein dynamics and progression-free survival for alternative isatuximab dosing with pomalidomide/dexamethasone. *Br. J. Clin. Pharmacol.* **88** 2052–2064.

van Buuren, S. (2021). *Flexible Imputation of Missing Data* 2nd ed. Chapman & Hall/CRC, Boca Raton, FL, USA.

van de Donk, N. W. C. J., Pawlyn, C. and Yong, K. L. (2021). Multiple myeloma. *Lancet* **397** 410–427.

Vehtari, A., Gelman, A. and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27** 1413–1432. MR3647105 https://doi.org/10.1007/s11222-016-9696-4

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. and Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Anal.* **16** 667–718. MR4298989 https://doi.org/10.1214/20-ba1221

Wang, C. Y., Wang, N. and Wang, S. (2000). Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. *Biometrics* **56** 487–495.

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

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44** 175–188. MR0931633 https://doi.org/10.2307/2531905

Zhou, G. C., Song, S. and Szczesniak, R. D. (2023). Multilevel joint model of longitudinal continuous and binary outcomes for hierarchically structured data. *Stat. Med.* **42** 2914–2927. MR4606377 https://doi.org/10.1002/sim.9758

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 322–345
https://doi.org/10.1214/25-AOAS2100

TEMPORAL MODELS FOR ESTIMATION AND SHORT-TERM FORECASTING OF NEONATAL MORTALITY RATES IN SUB-SAHARAN AFRICA

BY KATHERINE R. PAULSON[1,a], GEIR-ARNE FUGLSTAD[2,b], ZEHANG RICHARD LI[3,c] AND JONATHAN WAKEFIELD[4,d]

[1]*Department of Biostatistics, University of Washington,* [a]*krpaul@uw.edu*
[2]*Department of Mathematical Sciences, Norwegian University of Science and Technology,* [b]*geir-arne.fuglstad@ntnu.no*
[3]*Department of Statistics, University of California Santa Cruz,* [c]*lizehang@ucsc.edu*
[4]*Departments of Biostatistics and Statistics, University of Washington,* [d]*jonno@uw.edu*

Accurate estimation and forecasts for neonatal mortality rates (NMRs) in low- and middle-income countries is an urgent problem. Much of child mortality is preventable, and understanding temporal trends is of great interest when evaluating past performance and planning future policy or programming. In countries without robust vital registration, we rely on modeled estimates based on survey data to understand trends. A toolkit of compelling temporal models exists, but these methods have not been comprehensively evaluated for their application for the estimation of the NMR in low- and middle-income countries using household survey data. Using Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS) data from 41 countries in sub-Saharan Africa, we estimate and forecast the national-level NMR for 1970–2030 separately with random walk, auto-regressive, penalized spline, natural spline, and logit-linear latent temporal models. We examine the statistical behavior of these temporal models with both an out-of-sample analysis using the DHS and MICS data and a simulation study. We find that the second-order random walk and the penalized spline have the least bias, and short-term forecasts from the penalized spline tend to have narrower intervals with better out-of-sample performance. From the analysis of the NMR in sub-Saharan Africa, we estimate that six or fewer of the 41 countries included are on track to achieve the Sustainable Development Goals target of 12 neonatal deaths per 1000 live births by 2030.

REFERENCES

ADIN, A., KRAINSKI, E. T., LENZI, A., LIU, Z., MARTÍNEZ-MINAYA, J. and RUE, H. (2024). Automatic cross-validation in structured models: Is it time to leave out leave-one-out? *Spat. Stat.* **62** Paper No. 100843, 17. MR4760950 https://doi.org/10.1016/j.spasta.2024.100843

ALEXANDER, M. and ALKEMA, L. (2018). Global estimation of neonatal mortality using a Bayesian hierarchical splines regression model. *Demogr. Res.* **38** 335–372.

ALKEMA, L. and NEW, J. R. (2014). Global estimation of child mortality using a Bayesian B-spline Bias-reduction model. *Ann. Appl. Stat.* **8** 2122–2149. MR3292491 https://doi.org/10.1214/14-AOAS768

BREIDT, F. J. and OPSOMER, J. D. (2017). Model-assisted survey estimation with modern prediction techniques. *Statist. Sci.* **32** 190–205. MR3648955 https://doi.org/10.1214/16-STS589

BÜRKNER, P.-C., GABRY, J. and VEHTARI, A. (2020). Approximate leave-future-out cross-validation for Bayesian time series models. *J. Stat. Comput. Simul.* **90** 2499–2523. MR4145352 https://doi.org/10.1080/00949655.2020.1783262

CROFT, T. N., ALLEN, C. K., ZACHARY, B. W. et al. (2023). Guide to DHS Statistics. ICF, Rockville, Maryland, USA.

EILERS, P. H. C. and MARX, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statist. Sci.* **11** 89–121. With comments and a rejoinder by the authors. MR1435485 https://doi.org/10.1214/ss/1038425655

Key words and phrases. Random walk, penalized spline, autoregressive, temporal latent effects, Bayesian, neonatal mortality rate.

FEDERAL MINISTRY OF HEALTH AND SOCIAL WELFARE OF NIGERIA (FMOHSW), NATIONAL POPULATION COMMISSION (NPC) [NIGERIA] and ICF (2024). Nigeria Demographic and Health Survey 2023-24: Key Indicators Report. Abuja, Nigeria, and Rockville, Maryland, USA: NPC and ICF.

FINLAY, J. E., ÖZALTIN, E. and CANNING, D. (2011). The association of maternal age with infant mortality, child anthropometric failure, diarrhoea and anaemia for first births: Evidence from 55 low- and middle-income countries. *BMJ Open* **1** e000226.

FOREMAN, K. J., MARQUEZ, N., DOLGERT, A., FUKUTAKI, K., FULLMAN, N., MCGAUGHEY, M., PLETCHER, M. A., SMITH, A. E., TANG, K. et al. (2018). Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: Reference and alternative scenarios for 2016–40 for 195 countries and territories. *Lancet* **392** 2052–2090.

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

GODWIN, J. and WAKEFIELD, J. (2021). Space-time modeling of child mortality at the Admin-2 level in a low and middle income countries context. *Stat. Med.* **40** 1593–1638. MR4229793 https://doi.org/10.1002/sim.8854

GÓMEZ-RUBIO, V. (2020). *Bayesian Inference with INLA*. CRC Press, Boca Raton, FL.

HÁJEK, J. (1971). Discussion of "An essay on the logical foundations of survey sampling, part I," by D. Basu. In *Foundations of Statistical Inference* (*Proc. Sympos.*, *Univ. Waterloo*, *Waterloo*, *Ont.*, 1970). Holt, Rinehart & Winston, Toronto.

HUG, L., ALEXANDER, M., YOU, D., ALKEMA, L. and UN INTER-AGENCY GROUP FOR CHILD MORTALITY ESTIMATION (2019). National, regional, and global levels and trends in neonatal mortality between 1990 and 2017, with scenario-based projections to 2030: A systematic analysis. *Lancet Glob. Health* **7** e710–e720.

KRISTENSEN, K., NIELSEN, A., BERG, C. W., SKAUG, H. and BELL, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *J. Stat. Softw.* **70** 1–21.

LANG, S. and BREZGER, A. (2004). Bayesian P-splines. *J. Comput. Graph. Statist.* **13** 183–212. MR2044877 https://doi.org/10.1198/1061860043010

LEE, R. D. and CARTER, L. R. (1992). Modeling and forecasting U.S. mortality. *J. Amer. Statist. Assoc.* **87** 659–671.

LIU, Z. and RUE, H. (2022). Leave-group-out cross-validation for latent Gaussian models.

LUMLEY, T. (2010). *Complex Surveys*: *A Guide to Analysis Using R*: *A Guide to Analysis Using R*. Wiley, New York.

LUMLEY, T. (2023). survey: Analysis of complex survey samples. R package version 4.2.

MAHY, M. (2003). Measuring child mortality in AIDS-affected countries. Technical Report, United Nations, New York.

MERCER, L. D., WAKEFIELD, J., PANTAZIS, A., LUTAMBI, A. M., MASANJA, H. and CLARK, S. (2015). Space-time smoothing of complex survey data: Small area estimation for child mortality. *Ann. Appl. Stat.* **9** 1889–1905. MR3456357 https://doi.org/10.1214/15-AOAS872

MSEMBURI, W., KARLINSKY, A., KNUTSON, V., ALESHIN-GUENDEL, S., CHATTERJI, S. and WAKEFIELD, J. (2023). The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature* **613** 130–137.

NOORI, N., PROCTOR, J. L., EFEVBERA, Y. and ORON, A. P. (2022). The effect of adolescent pregnancy on child mortality in 46 low- and middle-income countries. *BMJ Glob. Health.* **7** e007681.

PAULSON, K. R., FUGLSTAD, G.-A., LI, Z. R. and WAKEFIELD, J. (2026). Supplement to "Temporal models for estimation and short-term forecasting of neonatal mortality rates in sub-Saharan Africa." https://doi.org/10.1214/25-AOAS2100SUPP

PAULSON, K. R., KAMATH, A. M., ALAM, T., BIENHOFF, K., HAY, S. I., MURRAY, C. J. L., WANG, H., KASSEBAUM, N. J. et al. (2021). Global, regional, and national progress towards Sustainable Development Goal 3.2 for neonatal and child health: All-cause and cause-specific mortality findings from the Global Burden of Disease Study 2019. *Lancet* **398** 870–905.

PEDERSEN, J. and LIU, J. (2012). Child mortality estimation: Appropriate time periods for child mortality estimates from full birth histories. *PLoS Med.* **9** e1001289.

PRIETO, J. R., VERHULST, A. and GUILLOT, M. (2021). Estimating the infant mortality rate from DHS birth histories in the presence of age heaping. *PLoS ONE* **16** e0259304.

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

RUE, H., RIEBLER, A., SØRBYE, S. H., ILLIAN, J. B., SIMPSON, D. P. and LINDGREN, F. K. (2017). Bayesian computing with INLA: A review. *Annu. Rev. Stat. Appl.* **4** 395–421.

SCHUMACHER, A. E., KYU, H. H., LIM, S. S., MURRAY, C. J. L. et al. (2024). Global age-sex-specific mortality, life expectancy, and population estimates in 204 countries and territories and 811 subnational locations, 1950–2021, and the impact of the COVID-19 pandemic: A comprehensive demographic analysis for the Global Burden of Disease Study 2021. *Lancet* **403** 1989–2056.

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

SØRBYE, S. H. and RUE, H. (2014). Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spat. Stat.* **8** 39–51. MR3326820 https://doi.org/10.1016/j.spasta.2013.06.004

SUSMANN, H., ALEXANDER, M. and ALKEMA, L. (2022). Temporal models for demographic and global health outcomes in multiple populations: Introducing a new framework to review and standardise documentation of model assumptions and facilitate model comparison. *Int. Stat. Rev.* **90** 437–467. MR4524819 https://doi.org/10.1111/insr.12491

UNITED NATIONS CHILDREN'S FUND (UNICEF) (2025). Levels & Trends in Child Mortality.

UNITED NATIONS INTER-AGENCY GROUP FOR CHILD MORTALITY ESTIMATION (UN IGME) (2024). Explanatory Notes: Child, adolescent and youth mortality trend series to 2023.

VENTRUCCI, M. and RUE, H. (2016). Penalized complexity priors for degrees of freedom in Bayesian P-splines. *Stat. Model.* **16** 429–453. MR3589051 https://doi.org/10.1177/1471082X16659154

WAKEFIELD, J., FUGLSTAD, G.-A., RIEBLER, A., GODWIN, J., WILSON, K. and CLARK, S. J. (2019). Estimating under-five mortality in space and time in a developing world context. *Stat. Methods Med. Res.* **28** 2614–2634. MR4000184 https://doi.org/10.1177/0962280218767988

WALKER, N., HILL, K. and ZHAO, F. (2012). Child mortality estimation: Methods used to adjust for bias due to AIDS in estimating trends in under-five mortality. *PLoS Med.* **9** e1001298.

WILSON, K. and WAKEFIELD, J. (2021). Child mortality estimation incorporating summary birth history data. *Biometrics* **77** 1456–1466. MR4357851 https://doi.org/10.1111/biom.13383

WORLD HEALTH ORGANIZATION (WHO) and UNITED NATIONS CHILDREN'S FUND (UNICEF) (2020). Ending preventable newborn deaths and stillbirths by 2030. Available at https://www.unicef.org/reports/ending-preventable-newborn-deaths-stillbirths-quality-health-coverage-2020-2025.

WORLD HEALTH ORGANIZATION (WHO) (2024). Newborn mortality fact sheet. Available at https://www.who.int/news-room/fact-sheets/detail/newborn-mortality.

WU, H., ZHAO, M., LIANG, Y., LIU, F. and XI, B. (2021). Maternal age at birth and neonatal mortality: Associations from 67 low-income and middle-income countries. *Paediatr. Perinat. Epidemiol.* **35** 318–327.

YAO, Y., VEHTARI, A., SIMPSON, D. and GELMAN, A. (2018). Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Anal.* **13** 917–1007. Including a rejoinder by the authors. MR3853125 https://doi.org/10.1214/17-BA1091

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 346–363
https://doi.org/10.1214/25-AOAS2111

LATENT CLASS ANALYSIS WITH DISCRETE FAILURE TIME MODEL

BY QINMENGGE LI[1,a], KEVIN HE[1,b], LAM C. TSOI[1,2,3,4,c] AND JIAN KANG[1,d]

[1]*Department of Biostatistics, University of Michigan,* [a]*liqinmg@umich.edu,* [b]*kevinhe@umich.edu,* [c]*alextsoi@umich.edu,* [d]*jiankang@umich.edu*

[2]*Department of Computational Medicine and Bioinformatics, University of Michigan Medical School*

[3]*Department of Dermatology, University of Michigan Medical School*

[4]*Mary H. Weiser Food Allergy Center, University of Michigan*

In survival analysis, accurate identification of latent classes is essential to effectively account for potential hidden population heterogeneity. In response to this challenge, we introduce the latent class discrete survival (LaCDS) model. LaCDS employs a finite-mixture model structure within the context of the discrete failure time model and implements the expectation-maximization algorithm for efficient optimization. Through extensive simulation studies, we evaluate the performance of LaCDS in comparison to other methods. Our results demonstrate LaCDS's superior ability to identify population heterogeneities, both in terms of baseline hazards and coefficients. Additionally, it is robust under both discrete and continuous simulation mechanisms. We apply LaCDS and other methods to identify subgroups among kidney transplant patients within the Organ Procurement and Transplantation Network (OPTN) study. Our findings underscore the superior accuracy of LaCDS in subgrouping homogeneous patients compared to existing methods.

REFERENCES

AKAIKE, H. (1974). A new look at the statistical model identification: System identification and time-series analysis. *IEEE Trans. Automat. Control* **AC-19** 716–723. MR0423716 https://doi.org/10.1109/tac.1974.1100705

ARSHAD, A., HODSON, J., CHAPPELOW, I., INSTON, N. G., READY, A. R., NATH, J. and SHARIF, A. (2018). The impact of donor body mass index on outcomes after deceased kidney transplantation–a national population-cohort study. *Transpl. Int.* **31** 1099–1109.

BLOOM, R. D. and AUGUSTINE, J. J. (2021). Beyond the biopsy: Monitoring immune status in kidney recipients. *Clin. J. Amer. Soc. Nephrol.* **16** 1413–1422.

BUČAR, T., NAGODE, M. and FAJDIGA, M. (2004). Reliability approximation using finite Weibull mixture distributions. *Reliab. Eng. Syst. Saf.* **84** 241–251.

DE ANGELIS, R., CAPOCACCIA, R., HAKULINEN, T., SODERMAN, B. and VERDECCHIA, A. (1999). Mixture models for cancer survival analysis: Application to population-based data with covariates. *Stat. Med.* **18** 441–454.

DO, D., DO, L. and NGUYEN, X. (2025). Strong identifiability and parameter learning in regression with heterogeneous response. *Electron. J. Stat.* **19** 131–203. MR4849921 https://doi.org/10.1214/24-ejs2339

DUBOURG, L., COCHAT, P., HADJ-AÏSSA, A., TYDÉN, G. and BERG, U. B. (2002). Better long-term functional adaptation to the child's size with pediatric compared to adult kidney donors. *Kidney Int.* **62** 1454–1460.

EGLESTON, B. L., UZZO, R. G. and WONG, Y.-N. (2017). Latent class survival models linked by principal stratification to investigate heterogenous survival subgroups among individuals with early-stage kidney cancer. *J. Amer. Statist. Assoc.* **112** 534–546. MR3671750 https://doi.org/10.1080/01621459.2016.1240078

ENG, K. H. and HANLON, B. M. (2014). Discrete mixture modeling to address genetic heterogeneity in time-to-event regression. *Bioinformatics* **30** 1690–1697.

ERIŞOĞLU, U., ERIŞOĞLU, M. and EROL, H. (2011). A mixture model of two different distributions approach to the analysis of heterogeneous survival data. *Int. J. Comput. Math. Sci.* **5** 75–89.

FEI, T., HANFELT, J. and PENG, L. (2022). Latent class analysis with semi-parametric proportional hazards submodel for time-to-event data. arXiv Preprint. Available at arXiv:2202.00775.

GEDDES, C. C., CHURCH, C. C., COLLIDGE, T., MCCRUDEN, E. A., GILLESPIE, G., MATTHEWS, E., HAINMUELLER, A. and BRIGGS, J. D. (2003). Management of cytomegalovirus infection by weekly surveillance after renal transplant: Analysis of cost, rejection and renal function. *Nephrol. Dial. Transplant.* **18** 1891–1898.

Key words and phrases. Survival analysis, latent class, finite mixture model.

GRASSMANN, A., GIOBERGE, S., MOELLER, S. and BROWN, G. (2005). ESRD patients in 2004: Global overview of patient numbers, treatment modalities and associated trends. *Nephrol. Dial. Transplant.* **20** 2587–2593.

GREENHOUSE, J. and SILLIMAN, N. (1996). Applications of a mixture survival model with covariates to the analysis of a depression prevention trial. *Stat. Med.* **15** 2077–2094.

GUPTA, A., CHEN, G. and KAPLAN, B. (2014). KDPI and donor selection. *Amer. J. Transplant.* **14** 2444–2445.

HART, A., SMITH, J., SKEANS, M., GUSTAFSON, S., WILK, A., ROBINSON, A., WAINRIGHT, J., HAYNES, C., SNYDER, J. et al. (2018). OPTN/SRTR 2016 annual data report: Kidney. *Amer. J. Transplant.* **18** 18–113.

HELD, P. J., MCCORMICK, F., OJO, A. and ROBERTS, J. P. (2016). A cost-benefit analysis of government compensation of kidney donors. *Amer. J. Transplant.* **16** 877–885.

HILTON, R. P., ZHENG, Y. and SERBAN, N. (2018). Modeling heterogeneity in healthcare utilization using massive medical claims data. *J. Amer. Statist. Assoc.* **113** 111–121. MR3803443 https://doi.org/10.1080/01621459.2017.1330203

HUMAR, A. and MATAS, A. J. (2005). Surgical complications after kidney transplantation. *Sem. Dial.* **18** 505–510.

HUNSBERGER, S., ALBERT, P. S. and LONDON, W. B. (2009). A finite mixture survival model to characterize risk groups of neuroblastoma. *Stat. Med.* **28** 1301–1314. MR2662180 https://doi.org/10.1002/sim.3543

KABORÉ, R., HALLER, M. C., HARAMBAT, J., HEINZE, G. and LEFFONDRÉ, K. (2017). Risk prediction models for graft failure in kidney transplantation: A systematic review. *Nephrol. Dial. Transplant.* **32** ii68–ii76.

KALBFLEISCH, J. D. and PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1924807 https://doi.org/10.1002/9781118032985

KARIM, A., FARRUGIA, D., CHESHIRE, J., MAHBOOB, S., BEGAJ, I., RAY, D. and SHARIF, A. (2014). Recipient age and risk for mortality after kidney transplantation in England. *Transplantation* **97** 832–838.

KUK, A. Y. and CHEN, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79** 531–541.

LAMBERT, P. C., DICKMAN, P. W., WESTON, C. L. and THOMPSON, J. R. (2010). Estimating the cure fraction in population-based cancer studies by using finite mixture models. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **59** 35–55. MR2750131 https://doi.org/10.1111/j.1467-9876.2009.00677.x

LENTINE, K. L., SMITH, J. M., MILLER, J. M., BRADBROOK, K., LARKIN, L., WEISS, S., HANDAROVA, D. K., TEMPLE, K., ISRANI, A. K. et al. (2023). OPTN/SRTR 2021 annual data report: Kidney. *Amer. J. Transplant.* **23** S21–S120.

LI, Q., HE, K., TSOI, L. C. and KANG, J. (2026). Supplement to “Latent Class Analysis with Discrete Failure Time Model.” https://doi.org/10.1214/25-AOAS2111SUPPA, https://doi.org/10.1214/25-AOAS2111SUPPB

LI, Y., TIWARI, R. C. and GUHA, S. (2007). Mixture cure survival models with dependent censoring. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 285–306. MR2323754 https://doi.org/10.1111/j.1467-9868.2007.00589.x

LYSAGHT, M. J. (2002). Maintenance dialysis population dynamics: Current trends and long-term implications. *J. Amer. Soc. Nephrol.* **13** S37–S40.

MAIR, P. and HUDEC, M. (2009). Multivariate Weibull mixtures with proportional hazard restrictions for dwell-time-based session clustering with incomplete data. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **58** 619–639. MR2750259 https://doi.org/10.1111/j.1467-9876.2009.00665.x

MATAS, A., SMITH, J., SKEANS, M., LAMB, K., GUSTAFSON, S., SAMANA, C., STEWART, D., SNYDER, J., ISRANI, A. et al. (2013). OPTN/SRTR 2011 annual data report: Kidney. *Amer. J. Transplant.* **13** 11–46.

MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models. Wiley Series in Probability and Statistics: Applied Probability and Statistics*. Wiley Interscience, New York. MR1789474 https://doi.org/10.1002/0471721182

MCLACHLAN, G. J., LEE, S. X. and RATHNAYAKE, S. I. (2019). Finite mixture models. *Annu. Rev. Stat. Appl.* **6** 355–378. MR3939525 https://doi.org/10.1146/annurev-statistics-031017-100325

MERION, R. M., ASHBY, V. B., WOLFE, R. A., DISTANT, D. A., HULBERT-SHEARON, T. E., METZGER, R. A., OJO, A. O. and PORT, F. K. (2005). Deceased-donor characteristics and the survival benefit of kidney transplantation. *J. Amer. Med. Assoc.* **294** 2726–2733.

MOELLER, S., GIOBERGE, S. and BROWN, G. (2002). ESRD patients in 2001: Global overview of patients, treatment modalities and development trends. *Nephrol. Dial. Transplant.* **17** 2071–2076.

MUTHÉN, B. and MASYN, K. (2005). Discrete-time survival mixture analysis. *J. Educ. Behav. Stat.* **30** 27–58.

MUTHÉN, B. and SHEDDEN, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* **55** 463–469.

RAO, P. S., SCHAUBEL, D. E., JIA, X., LI, S., PORT, F. K. and SARAN, R. (2007). Survival on dialysis post–kidney transplant failure: Results from the scientific registry of transplant recipients. *Amer. J. Kidney Dis.* **49** 294–300.

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.

SALIFU, M. O., TEDLA, F. and MARKELL, M. S. (2005). Management of the well renal transplant recipient: Outpatient surveillance and treatment recommendations. In *Seminars in Dialysis* **18** 520–528. Wiley Online Library.

SARAN, R., ROBINSON, B., ABBOTT, K. C., AGODOA, L. Y., BHAVE, N., BRAGG-GRESHAM, J., BALKRISHNAN, R., DIETRICH, X., ECKARD, A. et al. (2018). US renal data system 2017 annual data report: Epidemiology of kidney disease in the United States. *Amer. J. Kidney Dis*. **71** A7.

SARWAL, M., CHUA, M.-S., KAMBHAM, N., HSIEH, S.-C., SATTERWHITE, T., MASEK, M. and SALVATIERRA JR, O. (2003). Molecular heterogeneity in acute renal allograft rejection identified by DNA microarray profiling. *N. Engl. J. Med*. **349** 125–138.

SCHNITZLER, M. A., LENTINE, K. L., AXELROD, D., GHEORGHIAN, A., YOU, M., KALSEKAR, A. and L'ITALIEN, G. (2012). Use of 12-month renal function and baseline clinical factors to predict long-term graft survival: Application to BENEFIT and BENEFIT-EXT trials. *Transplantation* **93** 172–181.

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

SNYDER, J. J., SALKOWSKI, N., KIM, S. J., ZAUN, D., XIONG, H., ISRANI, A. K. and KASISKE, B. L. (2016). Developing statistical models to assess transplant outcomes using national registries: The process in the United States. *Transplantation* **100** 288–294.

TERASAKI, P. I., GJERTSON, D. W., CECKA, J. M., TAKEMOTO, S. and CHO, Y. W. (1997). Significance of the donor age effect on kidney transplants. *Clin. Transplant*. **11** 366–372.

VAN DER LEEST, R. J., ZACHOW, K. R., OSTROW, R. S., BENDER, M., PASS, F. and FARAS, A. J. (1987). Human papillomavirus heterogeneity in 36 renal transplant recipients. *Arch. Dermatol*. **123** 354–357.

VEROUX, M., GROSSO, G., CORONA, D., MISTRETTA, A., GIAQUINTA, A., GIUFFRIDA, G., SINAGRA, N. and VEROUX, P. (2012). Age is an important predictor of kidney transplantation outcome. *Nephrol. Dial. Transplant*. **27** 1663–1671.

WU, R.-F., ZHENG, M. and YU, W. (2016). Subgroup analysis with time-to-event data under a logistic-Cox mixture model. *Scand. J. Stat*. **43** 863–878. MR3543327 https://doi.org/10.1111/sjos.12213

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 364–384
https://doi.org/10.1214/25-AOAS2121

LATENT SPACE MODELING FOR HUMAN DISEASE NETWORK WITH TEMPORAL VARIATIONS: ANALYSIS OF MEDICARE DATA

BY GUOJUN ZHU[1,a], RUIYUE WANG[1,b], RONG LI[2,d], SANGUO ZHANG[1,c], SHUANGGE MA[2,e], GUANZHONG QIAO[3,f] AND HAO MEI[4,g]

[1]*School of Mathematical Sciences, University of Chinese Academy of Sciences,* [a]*zhuguojun23@mails.ucas.ac.cn,* [b]*wangruiyue21@mails.ucas.ac.cn,* [c]*sgzhang@ucas.ac.cn*

[2]*Department of Biostatistics, Yale School of Public Health,* [d]*rong.li.rl946@yale.edu,* [e]*shuangge.ma@yale.edu*

[3]*Department of Orthopaedic, The First Hospital of Tsinghua University,* [f]*qgzh8916@163.com*

[4]*Center for Applied Statistics, School of Statistics, Institute of Health Data Science, Renmin University of China,* [g]*hao.mei@ruc.edu.cn*

Human disease network (HDN) analysis, which jointly considers a large number of diseases and focuses on their interconnections, is getting increasingly popular and can shed important insight not possessed by individual-disease-based analysis. Multiple network analysis techniques have been developed for HDNs, although new developments are still strongly needed. In this article we adopt latent space modeling, which has proven powerful in other network analysis contexts and offers unique, insightful interpretations, but has been limitedly applied in HDN analysis. Different from some other types of network analysis and some other HDN analyses (such as gene-centric ones), in this article we pay unique attention to modeling temporal variations. For this purpose, a penalization approach is developed, which can identify time regions with constant network structures (that correspond to ignorable changes) as well as those with smooth variations. The statistical and computational properties are rigorously established. With Medicare data—one of the most powerful medical claims databases—we analyze the admission records of 133 million hospital inpatient treatments from January 2008 to December 2019. Sensible findings are made on disease interconnections and clustering structures. Additionally, the temporal variations, which have not been revealed in the literature, are found to be interpretable. The analysis can provide a new way for connecting and grouping diseases and assist in understanding and planning medical resources.

REFERENCES

ABRAHAM, G., KOWALCZYK, A., ZOBEL, J. and INOUYE, M. (2013). Performance and robustness of penalized and unpenalized methods for genetic prediction of complex human disease. *Genet. Epidemiol.* **37** 184–195.

AUBERT, C. E., SCHNIPPER, J. L., FANKHAUSER, N., MARQUES-VIDAL, P., STIRNEMANN, J., AUERBACH, A. D., ZIMLICHMAN, E., KRIPALANI, S., VASILEVSKIS, E. E. et al. (2020). Association of patterns of multimorbidity with length of stay: A multinational observational study. *Medicine* **99** e21650.

BARABÁSI, A.-L., GULBAHCE, N. and LOSCALZO, J. (2011). Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12** 56–68.

CARSON, J. L., TERRIN, M. L., NOVECK, H., SANDERS, D. W., CHAITMAN, B. R., RHOADS, G. G., NEMO, G., DRAGERT, K., BEAUPRE, L. et al. (2011). Liberal or restrictive transfusion in high-risk patients after hip surgery. *N. Engl. J. Med.* **365** 2453–2462.

COTTERILL, P. G. (2023). An assessment of completeness and medical coding of Medicare Advantage hospitalizations in two national data sets. *Health Serv. Res.* **58** 1303–1313.

CROWSON, C. S., GUNDERSON, T. M., DYKHOFF, H. J., MYASOEDOVA, E., ATKINSON, E. J., KRONZER, V. L., COFFEY, C. M. and DAVIS, J. M. III (2022). Comprehensive assessment of multimorbidity burden in a population-based cohort of patients with rheumatoid arthritis. *RMD Open* **8** e002022.

Key words and phrases. Human disease network, latent space modeling, temporal variation, penalized estimation, Medicare data.

Davis, D. A. and Chawla, N. V. (2011). Exploring and exploiting disease interactions from multi-relational gene and phenotype networks. *PLoS ONE* **6** e22670.

Fung, V., Brand, R., Newhouse, J. and Hsu, J. (2011). Using medicare data for comparative effectiveness research–opportunities and challenges. *Amer. J. Manag. Care* **17** 488.

Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M. and Barabási, A.-L. (2007). The human disease network. *Proc. Natl. Acad. Sci. USA* **104** 8685–8690.

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

Holland, P. W., Laskey, K. B. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M. P. et al. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature* **545** 505–509.

Ideker, T. and Krogan, N. J. (2012). Differential network biology. *Mol. Syst. Biol.* **8** 565.

Jensen, E. T., Cook, S. F., Allen, J. K., Logie, J., Brookhart, M. A., Kappelman, M. D. and Dellon, E. S. (2015). Enrollment factors and bias of disease prevalence estimates in administrative claims data. *Ann. Epidemiol.* **25** 519–525.

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** 4557.

Jin, J., Ke, Z. T. and Luo, S. (2024). Mixed membership estimation for social networks. *J. Econometrics* **239** Paper No. 105369, 17. MR4708611 https://doi.org/10.1016/j.jeconom.2022.12.003

Kim, B., Lee, K. H., Xue, L. and Niu, X. (2018). A review of dynamic network models with latent variables. *Stat. Surv.* **12** 105–135. MR3850294 https://doi.org/10.1214/18-SS121

Liu, M., Fan, X. and Ma, S. (2024). A quantitative linguistic analysis of a cancer online health community with a smooth latent space model. *Ann. Appl. Stat.* **18** 144–158. MR4698602 https://doi.org/10.1214/23-aoas1783

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, 67. MR4071187

Martin, A. B., Hartman, M., Washington, B., Catlin, A. and National Health Expenditure Accounts Team (2025). National health expenditures in 2023: Faster growth as insurance coverage and utilization increased: Article examines national health expenditures in 2023. *Health Aff.* **44** 12–22.

Martinez, R., Morsch, P., Soliz, P., Hommes, C., Ordunez, P. and Vega, E. (2021). Life expectancy, healthy life expectancy, and burden of disease in older people in the Americas, 1990–2019: A population-based study. *Rev. Panamer. Salud Pública* **45** e114.

Mei, H., Jia, R., Qiao, G., Lin, Z. and Ma, S. (2021). Human disease clinical treatment network for the elderly: The analysis of medicare inpatient length of stay data. *Stat. Med.* **40** 2083–2099. MR4229840 https://doi.org/10.1002/sim.8893

Mei, H., Jia, R., Qiao, G., Lin, Z. and Ma, S. (2023). Human disease clinical treatment network for the elderly: Analysis of the medicare inpatient length of stay and readmission data. *Biometrics* **79** 404–416. MR4572531 https://doi.org/10.1111/biom.13549

Mei, H., Wang, Z., Yang, H., Li, X. and Xu, Y. (2025a). Network analysis of multivariate time series data in biological systems: Methods and applications. *Brief. Bioinform.* **26** bbaf223.

Mei, H., Xiao, H., Shia, B.-C., Qiao, G. and Li, Y. (2025b). Interconnections of multimorbidity-related clinical outcomes: Analysis of health administrative claims data with a dynamic network approach. *Stat. Med.* **44** Paper No. e70125, 17. MR4907741 https://doi.org/10.1002/sim.70125

Milano, W. and Capasso, A. (2018). Neuroendocrine and metabolic disorders in bulimia nervosa. *Endocr. Metab. Immune Disord. Drug Targets* **18** 297–305.

Richards, A. L., Eckhardt, M. and Krogan, N. J. (2021). Mass spectrometry-based protein–protein interaction networks for the study of human diseases. *Mol. Syst. Biol.* **17** e8792.

Rosenkrantz, A. B., Hughes, D. R. and Duszak, R. Jr (2017). Medicare claims data resources: A primer for policy-focused radiology health services researchers. *J. Amer. Coll. Radiol.* **14** 1538–1544.

Salsabili, M., Kiogou, S. and Adam, T. J. (2020). The evaluation of clinical classifications software using the national inpatient sample database. *AMIA Summits Transl. Sci. Proc.* **2020** 542.

Schiltz, N. K., Warner, D. F., Sun, J., Bakaki, P. M., Dor, A., Given, C. W., Stange, K. C. and Koroukian, S. M. (2017). Identifying specific combinations of multimorbidity that contribute to health care resource utilization: An analytic approach. *Med. Care* **55** 276–284.

Shangguan, X., Xiong, J., Shi, S., Liao, Y., Chen, L., Deng, J., Wu, W., Wang, J., Tu, J. et al. (2022). Impact of the malnutrition on mortality in patients with osteoporosis: A cohort study from NHANES 2005-2010. *Front. Nutr.* **9** 868166.

Stone, K., Zwiggelaar, R., Jones, P. and Mac Parthaláin, N. (2022). A systematic review of the prediction of hospital length of stay: Towards a unified framework. *PLOS Digit. Health* **1** e0000017.

THAPI, S., BAEG, K., KIM, M. K. and GALLAGHER, E. J. (2021). Survival of patients with gastroenteropancreatic neuroendocrine tumors and diabetes mellitus. *Pancreas* **50** 1293–1297.

TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 91–108. MR2136641 https://doi.org/10.1111/j.1467-9868.2005.00490.x

TIBSHIRANI, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Ann. Statist.* **42** 285–323. MR3189487 https://doi.org/10.1214/13-AOS1189

WEI, W.-Q., BASTARACHE, L. A., CARROLL, R. J., MARLO, J. E., OSTERMAN, T. J., GAMAZON, E. R., COX, N. J., RODEN, D. M. and DENNY, J. C. (2017). Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS ONE* **12** e0175508.

WEN, Q., GAO, J., SONG, X., SUN, L. and TAN, J. (2019). Robusttrend: A Huber loss with a combined first and second order difference regularization for time series trend filtering. In *International Joint Conference on Artificial Intelligence* 3856–3862. Proceedings of Machine Learning Research.

YANG, P., QIU, H., WANG, L. and ZHOU, L. (2022). Early prediction of high-cost inpatients with ischemic heart disease using network analytics and machine learning. *Expert Syst. Appl.* **210** 118541.

ZHANG, M., CAI, B., LI, D., NIU, X. and ZHANG, J. (2024a). Preferential latent space models for networks with textual edges. Preprint. Available at arXiv:2405.15038.

ZHANG, X., XUE, S. and ZHU, J. (2020). A flexible latent space model for multilayer networks. In *International Conference on Machine Learning* 11288–11297. Proceedings of Machine Learning Research.

ZHANG, Y., PAN, R., ZHU, X., FANG, K. and WANG, H. (2025). A latent space model for weighted keyword co-occurrence networks with applications in knowledge discovery in statistics. *J. Comput. Graph. Statist.* **34** 779–794. MR4950361 https://doi.org/10.1080/10618600.2024.2407465

ZHANG, Y., ZHANG, J., SUN, Y. and WANG, J. (2024b). Change point detection in dynamic networks via regularized tensor decomposition. *J. Comput. Graph. Statist.* **33** 515–524. MR4754809 https://doi.org/10.1080/10618600.2023.2240864

ZHOU, D., WANG, L., DING, S., SHEN, M. and QIU, H. (2022). Phenotypic disease network analysis to identify comorbidity patterns in hospitalized patients with ischemic heart disease using large-scale administrative data. *Healthcare* **10** 80.

ZHU, G., WANG, R., LI, R., ZHANG, S., MA, S., QIAO, G. and MEI, H. (2026). Supplement to "Latent space modeling for human disease network with temporal variations: Analysis of medicare data." https://doi.org/10.1214/25-AOAS2121SUPPA, https://doi.org/10.1214/25-AOAS2121SUPPB

ZINMAN, B., WANNER, C., LACHIN, J. M., FITCHETT, D., BLUHMKI, E., HANTEL, S., MATTHEUS, M., DEVINS, T., JOHANSEN, O. E. et al. (2015). Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N. Engl. J. Med.* **373** 2117–2128.

NATIONAL COUNCIL ON AGING (2025). The Top 10 Most Common Chronic Conditions in Older Adults. https://www.ncoa.org/article/the-top-10-most-common-chronic-conditions-in-older-adults. Accessed: 2025-08-15.

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 385–407
https://doi.org/10.1214/25-AOAS2128

A DATA ENVELOPMENT ANALYSIS APPROACH FOR ASSESSING FAIRNESS IN RESOURCE ALLOCATION: APPLICATION TO KIDNEY EXCHANGE PROGRAMS

BY ALI KAAZEMPUR-MOFRAD[a] AND XIAOWU DAI[b]

[1]*Department of Statistics and Data Science, University of California, Los Angeles*, [a]*amofrad@ucla.edu*, [b]*daix@ucla.edu*

Kidney exchange programs have substantially increased transplantation rates but also raise critical concerns about fairness in organ allocation. We propose a novel framework leveraging Data Envelopment Analysis (DEA) to evaluate multiple dimensions of fairness—Priority, Access, and Outcome—within a unified model. This approach captures complexities often missed in single-metric analyses. Using data from the United Network for Organ Sharing, we separately quantify fairness across these dimensions: Priority Fairness through waitlist durations, Access Fairness via the Living Kidney Donor Profile Index (LKDPI) scores, and Outcome Fairness based on graft lifespan. We then apply our conditional DEA model with covariate adjustment to demonstrate significant disparities in kidney allocation efficiency across ethnic groups. To quantify uncertainty, we employ conformal prediction within a novel reference frontier mapping (RFM) framework, yielding group-conditional prediction intervals with finite-sample coverage guarantees. Our findings show notable differences in efficiency distributions between ethnic groups. Our study provides a rigorous framework for evaluating fairness in complex resource allocation systems with resource scarcity and mutual compatibility constraints.

REFERENCES

AGARWAL, N., ASHLAGI, I., AZEVEDO, E., FEATHERSTONE, C. R. and KARADUMAN, Ö. (2019). Market failure in kidney exchange. *Amer. Econ. Rev.* **109** 4026–4070.

AXELROD, D. A., SCHNITZLER, M. A., XIAO, H., IRISH, W., TUTTLE-NEWHALL, E., CHANG, S.-H., KASISKE, B. L., ALHAMAD, T. and LENTINE, K. L. (2018). An economic assessment of contemporary kidney transplant practice. *Amer. J. Transplant.* **18** 1168–1176.

BĂDIN, L., DARAIO, C. and SIMAR, L. (2012). How to measure the impact of environmental factors in a nonparametric production model. *European J. Oper. Res.* **223** 818–833.

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, Methodol.* **57** 289–300.

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

CENTERS FOR DISEASE CONTROL AND PREVENTION (2023). Chronic kidney disease in the United States, 2023. US Dept. Health and Human Services, Centers for Disease Control and Prevention, Atlanta, GA.

CHARNES, A., COOPER, W. W. and RHODES, E. (1978). Measuring the efficiency of decision making units. *European J. Oper. Res.* **2** 429–444.

CHILINGERIAN, J. A. (1995). Evaluating physician efficiency in hospitals: A multivariate analysis of best practices. *European J. Oper. Res.* **80** 548–574.

CHORPPATH, A. K. and ALPCAN, T. (2011). Learning user preferences in mechanism design. In 2011 50*th IEEE Conference on Decision and Control and European Control Conference* 5349–5355. IEEE.

COHEN, J. B., BLOOM, R. D., REESE, P. P., PORRETT, P. M., FORDE, K. A. and SAWINSKI, D. L. (2016). National outcomes of kidney transplantation from deceased diabetic donors. *Kidney Int.* **89** 636–647.

CORESH, J., SELVIN, E., STEVENS, L. A., MANZI, J., KUSEK, J. W., EGGERS, P., VAN LENTE, F. and LEVEY, A. S. (2007). Prevalence of chronic kidney disease in the United States. *J. Amer. Med. Assoc.* **298** 2038–2047.

Key words and phrases. Conformal prediction, data envelopment analysis, fairness, kidney exchange, resource allocation.

FAN, P.-Y., ASHBY, V. B., FULLER, D., BOULWARE, L., KAO, A., NORMAN, S. P., RANDALL, H., YOUNG, C., KALBFLEISCH, J. D. et al. (2010). Access and outcomes among minority transplant patients, 1999–2008, with a focus on determinants of kidney graft survival. *Amer. J. Transplant.* **10** 1090–1107.

FÄRE, R., GROSSKOPF, S. and LOVELL, C. K. (2013). *The Measurement of Efficiency of Production* 6. Springer, Berlin.

FINE, J. P. and GRAY, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *J. Amer. Statist. Assoc.* **94** 496–509.

GENTRY, S. E., MONTGOMERY, R. A. and SEGEV, D. L. (2011). Kidney paired donation: Fundamentals, limitations, and expansions. *Amer. J. Kidney Dis.* **57** 144–151.

GENTRY, S. E., SEGEV, D. L., SIMMERLING, M. and MONTGOMERY, R. A. (2007). Expanding kidney paired donation through participation by compatible pairs. *Amer. J. Transplant.* **7** 2361–2370.

GIBBS, I., CHERIAN, J. J. and CANDÈS, E. J. (2025). Conformal prediction with conditional guarantees. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **87** 1100–1126.

GILL, J. S., TONELLI, M., JOHNSON, N. and PEREIRA, B. J. (2004). Why do preemptive kidney transplant recipients have an allograft survival advantage? *Transplantation* **78** 873–879.

GORDON, E. J., MULLEE, J. O., RAMIREZ, D. I., MACLEAN, J., OLIVERO, M., FEINGLASS, J., CARNEY, P., O'CONNOR, K. and CAICEDO, J. C. (2014). Hispanic/Latino concerns about living kidney donation: A focus group study. *Prog. Transplant.* **24** 152–162.

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.

HARDT, M., PRICE, E. and SREBRO, N. (2016). Equality of opportunity in supervised learning. *Adv. Neural Inf. Process. Syst.* **29**.

HARIHARAN, S., ISRANI, A. K. and DANOVITCH, G. (2021). Long-term survival after kidney transplantation. *N. Engl. J. Med.* **385** 729–743.

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

IMAI, K., KEELE, L. and YAMAMOTO, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statist. Sci.* **25** 51–71.

KAAZEMPUR-MOFRAD, A. and DAI, X. (2026). Supplement to "A Data Envelopment Analysis Approach for Assessing Fairness in Resource Allocation: Application to Kidney Exchange Programs." https://doi.org/10.1214/25-AOAS2128SUPPA, https://doi.org/10.1214/25-AOAS2128SUPPB

KUCIRKA, L. M., GRAMS, M., BALHARA, K. S., JAAR, B. G. and SEGEV, D. L. (2012). Disparities in provision of transplant information affect access to kidney transplantation. *Amer. J. Transplant.* **12** 351–357.

KUSNER, M. J., LOFTUS, J., RUSSELL, C. and SILVA, R. (2017). Counterfactual fairness. *Adv. Neural Inf. Process. Syst.* **30**.

LENTINE, K., SMITH, J., HART, A., MILLER, J., SKEANS, M., LARKIN, L., ROBINSON, A., GAUNTT, K., ISRANI, A. et al. (2022). OPTN/SRTR 2020 annual data report: Kidney. *Amer. J. Transplant.* **22** 21–136.

MACKINNON, D. (2012). *Introduction to Statistical Mediation Analysis*. Routledge, London.

MALEK, S. K., KEYS, B. J., KUMAR, S., MILFORD, E. and TULLIUS, S. G. (2011). Racial and ethnic disparities in kidney transplantation. *Transpl. Int.* **24** 419–424.

MASSIE, A. B., GENTRY, S. E., MONTGOMERY, R. A., BINGAMAN, A. A. and SEGEV, D. L. (2013). Center-level utilization of kidney paired donation. *Amer. J. Transplant.* **13** 1317–1322.

MASSIE, A. B., LEANZA, J., FAHMY, L. M., CHOW, E. K., DESAI, N. M., LUO, X., KING, E. A., BOWRING, M. G. and SEGEV, D. L. (2016). A risk index for living donor kidney transplantation. *Amer. J. Transplant.* **16** 2077–2084.

MORENO-CALDERÓN, A., TONG, T. S. and THOKALA, P. (2020). Multi-criteria decision analysis software in healthcare priority setting: A systematic review. *PharmacoEconomics* **38** 269–283.

OGRYCZAK, W., LUSS, H., PIÓRO, M., NACE, D. and TOMASZEWSKI, A. (2014). Fair optimization and networks: A survey. *J. Appl. Math.* **2014** Art. ID 612018, 25.

OZCAN, Y. A. et al. (2008). *Health Care Benchmarking and Performance Evaluation*. Springer, Berlin.

PERSAD, G., WERTHEIMER, A. and EMANUEL, E. J. (2009). Principles for allocation of scarce medical interventions. *Lancet* **373** 423–431.

PINTO-RAMIREZ, J., GARCIA-LOPEZ, A., SALCEDO-HERRERA, S., PATINO-JARAMILLO, N., GARCIA-LOPEZ, J., BARBOSA-SALINAS, J., RIVEROS-ENRIQUEZ, S., HERNANDEZ-HERRERA, G. and GIRON-LUQUE, F. (2022). Risk factors for graft loss and death among kidney transplant recipients: A competing risk analysis. *PLoS ONE* **17** e0269990.

RAPAPORT, F. T. (1986). The case for a living emotionally related international kidney donor exchange registry. *Transplant. Proc.* **18** 5–9.

REES, M. A., DUNN, T. B., KUHR, C. S., MARSH, C. L., ROGERS, J., REES, S. E., CICERO, A., REECE, L. J., ROTH, A. E. et al. (2017). Kidney exchange to overcome financial barriers to kidney transplantation. *Amer. J. Transplant.* **17** 782–790.

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

SAKSHUWONG, S., ASHLAGI, I. and ROTH, A. E. (2023). StanfordKPD: a Kidney Exchange Platform for Kidney Paired Donation.

SHERO, J. A., AL OTAIBA, S., SCHATSCHNEIDER, C. and HART, S. A. (2022). Data envelopment analysis (DEA) in the educational sciences. *J. Exp. Educ.* **90** 1021–1040.

THOKALA, P., DEVLIN, N., MARSH, K., BALTUSSEN, R., BOYSEN, M., KALO, Z., LONGRENN, T., MUSSEN, F., PEACOCK, S. et al. (2016). Multiple criteria decision analysis for health care decision making—an introduction: Report 1 of the ISPOR MCDA emerging good practices task force. *Value Health* **19** 1–13.

TONELLI, M., WIEBE, N., KNOLL, G., BELLO, A., BROWNE, S., JADHAV, D., KLARENBACH, S. and GILL, J. (2011). Systematic review: Kidney transplantation compared with dialysis in clinically relevant outcomes. *Amer. J. Transplant.* **11** 2093–2109.

UNITED STATES RENAL DATA SYSTEM (2024). 2024 USRDS Annual Data Report: Epidemiology of kidney disease in the United States.

WANG, W., LEICHTMAN, A. B., REES, M. A., SONG, P. X.-K., ASHBY, V. B., SHEARON, T. and KALBFLEISCH, J. D. (2022). Kidney paired donation chains initiated by deceased donors. *Kidney Int. Rep.* **7** 1278–1288.

WANG, W., REES, M. A., LEICHTMAN, A. B., SONG, P. X.-K., BRAY, M., ASHBY, V. B., SHEARON, T., WHITEMAN, A. and KALBFLEISCH, J. D. (2021). Deceased donors as nondirected donors in kidney paired donation. *Amer. J. Transplant.* **21** 103–113.

WOLFE, R. A., ASHBY, V. B., MILFORD, E. L., OJO, A. O., ETTENGER, R. E., AGODOA, L. Y., 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.

WOLFE, R. A., MCCULLOUGH, K. P., SCHAUBEL, D. E., KALBFLEISCH, J. D., MURRAY, S., STEGALL, M. D. and LEICHTMAN, A. B. (2008). Calculating life years from transplant (LYFT): Methods for kidney and kidney-pancreas candidates. *Amer. J. Transplant.* **8** 997–1011.

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 408–433
https://doi.org/10.1214/25-AOAS2129
© Institute of Mathematical Statistics, 2026

ASYMPTOTICALLY EFFICIENT DATA-ADAPTIVE PENALIZED SHRINKAGE ESTIMATION WITH APPLICATION TO CAUSAL INFERENCE

BY HERBERT P. SUSMANN[1,a], YITING LI[2,d], MARA A. MCADAMS-DEMARCO[2,e], WENBO WU[1,b] AND IVÁN DÍAZ[1,c]

[1]*Division of Biostatistics, Department of Population Health, NYU Grossman School of Medicine,* [a]*susmah01@nyu.edu,* [b]*wenbowu@jhu.edu,* [c]*ivan.diaz@nyu.edu*

[2]*Department of Surgery, NYU Grossman School of Medicine,* [d]*yiting.li@nyulangone.org,* [e]*mara.mcadamsdemarco@nyulangone.org*

A rich literature exists on constructing nonparametric estimators with optimal asymptotic properties. In addition to asymptotic guarantees, it is often of interest to design estimators with desirable finite-sample properties, such as reduced mean-squared error of a large set of parameters. We provide examples drawn from causal inference where this may be the case, such as estimating a large number of group-specific treatment effects. We show how finite-sample properties of nonparametric estimators, particularly their variance, can be improved by careful application of *penalization*. Given a target parameter of interest, we derive a novel penalized parameter defined as the solution to an optimization problem that balances fidelity to the original parameter against a penalty term. By deriving the nonparametric efficiency bound for the penalized parameter, we are able to propose simple data-adaptive choices for the L_1 and L_2 tuning parameters designed to minimize finite-sample mean-squared error while preserving optimal asymptotic properties. The L_1 and L_2 penalization amounts to an adjustment that can be performed as a postprocessing step applied to any asymptotically normal and efficient estimator. We show in extensive simulations that this adjustment yields estimators with lower MSE than the unpenalized estimators. Finally, we apply our approach to estimate provider quality measures of kidney dialysis providers within a causal inference framework.

REFERENCES

ARMSTRONG, T. B., KOLESÁR, M. and PLAGBORG-MØLLER, M. (2022). Robust empirical Bayes confidence intervals. *Econometrica* **90** 2567–2602. MR4524894 https://doi.org/10.3982/ecta18597

BAHAMYIROU, A., SCHNITZER, M. E., KENNEDY, E. H., BLAIS, L. and YANG, Y. (2022). Doubly robust adaptive LASSO for effect modifier discovery. *The International Journal of Biostatistics* **18** 307–327. https://doi.org/10.1515/ijb-2020-0073

BENKESER, D. and VAN DER LAAN, M. (2016). The highly adaptive lasso estimator. In 2016 *IEEE International Conference on Data Science and Advanced Analytics* (*DSAA*) 689–696. https://doi.org/10.1109/DSAA.2016.93

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York. MR1623559

BRAKENHOFF, T. B., MOONS, K. G., KLUIN, J. and GROENWOLD, R. H. (2018). Investigating risk adjustment methods for health care provider profiling when observations are scarce or events rare. *Health Services Insights* **11** 1178632918785133. https://doi.org/10.1177/1178632918785133

DAIGNAULT, K. and SAARELA, O. (2017). Doubly robust estimator for indirectly standardized mortality ratios. *Epidemiologic Methods* **6** 20160016. https://doi.org/10.1515/em-2016-0016

DELONG, E. R., PETERSON, E. D., DELONG, D. M., MUHLBAIER, L. H., HACKETT, S. and MARK, D. B. (1997). Comparing risk-adjustment methods for provider profiling. *Stat. Med.* **16** 2645–2664. https://doi.org/10.1002/(SICI)1097-0258(19971215)16:23<2645::AID-SIM696>3.0.CO;2-D

DÍAZ, I. (2024). Non-agency interventions for causal mediation in the presence of intermediate confounding. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 435–460. MR4754091 https://doi.org/10.1093/jrsssb/qkad130

Key words and phrases. Causal inference, doubly robust estimation, penalization, shrinkage estimator.

EFRON, B. (2024). Empirical Bayes: Concepts and methods. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*. Chapman & Hall, London.

EFRON, B. and MORRIS, C. (1977). Stein's paradox in statistics. *Scientific American* **236** 119–127.

FELLER, A. and GELMAN, A. (2015). Hierarchical models for causal effects. In *Emerging Trends in the Social and Behavioral Sciences* 1-16 Wiley, New York. https://doi.org/10.1002/9781118900772.etrds0160

FRIEDMAN, J., TIBSHIRANI, R. and HASTIE, T. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33** 1–22. https://doi.org/10.18637/jss.v033.i01

HÁJEK, J. (1969/70). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* **14** 323–330. MR0283911 https://doi.org/10.1007/BF00533669

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California, Berkeley, Calif.*, 1970/1971), *Vol. I: Theory of Statistics* 175–194. Univ. California Press, Berkeley, CA. MR0400513

HANSEN, B. E. (2017). Stein-like 2SLS estimator. *Econometric Rev.* **36** 840–852. MR3680746 https://doi.org/10.1080/07474938.2017.1307579

HOERL, A. E., KANNARD, R. W. and BALDWIN, K. F. (1975). Ridge regression: Some simulations. *Communications in Statistics* **4** 105–123. https://doi.org/10.1080/03610927508827232

HOERL, A. E. and KENNARD, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12** 55–67. https://doi.org/10.1080/00401706.1970.10488634

IMDAD ULLAH, M., ASLAM, M. and ALTAF, S. (2018). Lmridge: A comprehensive R package for ridge regression. *The R Journal* **10** 326–346. https://doi.org/10.32614/RJ-2018-060

KALBFLEISCH, J. D. and HE, K. (2018). Discussion on "Time-dynamic profiling with application to hospital readmission among patients on dialysis," by Jason P. Estes, Danh V. Nguyen, Yanjun Chen, Lorien S. Dalrymple, Connie M. Rhee, Kamyar Kalantar-Zadeh, and Damla Senturk. *Biometrics* **74** 1401–1403.

KALBFLEISCH, J. D. and WOLFE, R. A. (2013). On monitoring outcomes of medical providers. *Statistics in Biosciences* **5** 286–302. https://doi.org/10.1007/s12561-013-9093-x

KAPLAN, D. M. and LIU, X. (2024). Confidence intervals for intentionally biased estimators. *Econometric Rev.* **43** 197–214. MR4720046 https://doi.org/10.1080/07474938.2024.2312288

KEIDING, N. and CLAYTON, D. (2014). Standardization and control for confounding in observational studies: A historical perspective. *Statist. Sci.* **29** 529–558. MR3300358 https://doi.org/10.1214/13-STS453

KENNEDY, E. H. (2016). Semiparametric theory and empirical processes in causal inference. In *Statistical Causal Inferences and Their Applications in Public Health Research. ICSA Book Ser. Stat.* 141–167. Springer, Cham. MR3617956

KENNEDY, E. H. (2024). Semiparametric doubly robust targeted double machine learning: A review. In *Handbook of Statistical Methods for Precision Medicine* 10 CRC Press, Boca Raton.

LE CAM, L. (1972). Limits of experiments. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California, Berkeley, Calif.*, 1970/1971), *Vol. I: Theory of Statistics* 245–261. Univ. California Press, Berkeley, CA. MR0415819

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

MAASOUMI, E. (1978). A modified Stein-like estimator for the reduced form coefficients of simultaneous equations. *Econometrica* **46** 695–703. MR0483203 https://doi.org/10.2307/1914241

MACKENZIE, T. A., GRUNKEMEIER, G. L., GRUNWALD, G. K., O'MALLEY, A. J., BOHN, C., WU, Y. and MALENKA, D. J. (2015). A primer on using shrinkage to compare in-hospital mortality between centers. *The Annals of Thoracic Surgery* **99** 757–761. https://doi.org/10.1016/j.athoracsur.2014.11.039

MCCLEAN, A., BRANSON, Z. and KENNEDY, E. H. (2024). Nonparametric estimation of conditional incremental effects. *J. Causal Inference* **12** Paper No. 20230024, 42. MR4736178 https://doi.org/10.1515/jci-2023-0024

NORMAND, S.-L. T., GLICKMAN, M. E. and GATSONIS, C. A. (1997). Statistical methods for profiling providers of medical care: Issues and applications. *J. Amer. Statist. Assoc.* **92** 803–814. https://doi.org/10.1080/01621459.1997.10474036

PFANZAGL, J. and WEFELMEYER, W. (1985). Contributions to a general asymptotic statistical theory. *Statistics & Risk Modeling* **3** 379–388.

SHORTREED, S. M. and ERTEFAIE, A. (2017). Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics* **73** 1111–1122. MR3744525 https://doi.org/10.1111/biom.12679

SMUCLER, E., ROTNITZKY, A. and ROBINS, J. M. (2019). A unifying approach for doubly-robust ℓ_1 regularized estimation of causal contrasts.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 197–206. Univ. California Press, Berkeley, CA. MR0084922

SUSMANN, H. and CHAMBAZ, A. (2023). Inference in Marginal Structural Models by Automatic Targeted Bayesian and Minimum Loss-Based Estimation.

SUSMANN, H., LI, Y., MCADAMS-DEMARCO, M. A., DÍAZ, I. and WU, W. (2025). Doubly Robust Nonparametric Efficient Estimation for Provider Evaluation. *J. R. Stat. Soc. Ser. A Stat. Soc.* MR2137327 https://doi.org/10.1093/jrsssa/qnaf145

SUSMANN, H. P, LI, Y., MCADAMS-DEMARCO, M. A, WU, W. and DÍAZ, I. (2026). Supplement to "Asymptotically efficient data-adaptive penalized shrinkage estimation with application to causal inference." https://doi.org/10.1214/25-AOAS2129SUPPA, https://doi.org/10.1214/25-AOAS2129SUPPB

TAY, J. K., NARASIMHAN, B. and HASTIE, T. (2023). Elastic net regularization paths for all generalized linear models. *Journal of Statistical Software* **106** 1–31. https://doi.org/10.18637/jss.v106.i01

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

TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data. Springer Series in Statistics*. Springer, New York. MR2233926

U.S. Renal Data System (2022). 2022 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.

VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statistics* **18** 309–348. MR0022330 https://doi.org/10.1214/aoms/1177730385

VAN DER LAAN, M. J. and RUBIN, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2** Art. 11, 40. MR2306500 https://doi.org/10.2202/1557-4679.1043

VAN DER VAART, A. W. (1992). Asymptotic linearity of minimax estimators. *Statist. Neerlandica* **46** 179–194. MR1178478 https://doi.org/10.1111/j.1467-9574.1992.tb01336.x

VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

VAREWYCK, M., GOETGHEBEUR, E., ERIKSSON, M. and VANSTEELANDT, S. (2014). On shrinkage and model extrapolation in the evaluation of clinical center performance. *Biostatistics* **15** 651–664. https://doi.org/10.1093/biostatistics/kxu019

WILLIAMSON, B. D., GILBERT, P. B., CARONE, M. and SIMON, N. (2021). Nonparametric variable importance assessment using machine learning techniques. *Biometrics* **77** 9–22. MR4229718 https://doi.org/10.1111/biom.13392

WU, W., YANG, Y., KANG, J. and HE, K. (2022). Improving large-scale estimation and inference for profiling health care providers. *Stat. Med.* **41** 2840–2853. MR4441590 https://doi.org/10.1002/sim.9387

YANG, X., PENG, B., CHEN, R., ZHANG, Q., ZHU, D., ZHANG, Q. J., XUE, F. and QI, L. (2014). Statistical profiling methods with hierarchical logistic regression for healthcare providers with binary outcomes. *J. Appl. Stat.* **41** 46–59. MR3291199 https://doi.org/10.1080/02664763.2013.830086

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 301–320. MR2137327 https://doi.org/10.1111/j.1467-9868.2005.00503.x

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 434–451
https://doi.org/10.1214/25-AOAS2130
© Institute of Mathematical Statistics, 2026

STRATIFIED REGRESSION ANALYSIS OF ZERO-TRUNCATED RECURRENT EVENT DATA

BY ANQI A. CHEN[1,2,a], X. JOAN HU[1,b] AND RHONDA J. ROSYCHUK[2,c]

[1] *Department of Statistics and Actuarial Science, Simon Fraser University,* [a]*aca142@sfu.ca,* [b]*joanh@stat.sfu.ca*
[2] *Department of Pediatrics, University of Alberta,* [c]*rhonda.rosychuk@ualberta.ca*

This paper is motivated by a pediatric mental health care (PMHC) program, which extracted the records of mental health-related emergency department (MHED) visits from population-based administrative databases during 2011–2017. Only information on the subjects with MHED visit experiences is available within a subject-specific time window. We focus on one of the program objectives: understanding how the visit occurrence is associated with the subject's past as well as their demographic and geographic exposures in the entire population. The available collection of the MHED records is framed as zero-truncated recurrent event data. We introduce an innovative stratified Cox regression model for the event process. The model is intensity-based but requires only a summary of the event history. We propose a new procedure for estimating the model parameters using the zero-truncated data integrated with some relevant population census information. We establish the consistency and asymptotic normality of the proposed estimator and examine its finite sample performance via extensive simulation in contrast with the maximum likelihood estimation based on the zero-truncated data only. The MHED data from the PMHC program are employed to illustrate the proposed approach throughout the paper.

REFERENCES

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. MR0673646

BRESLOW, N. E. (1972). Discussion of Professor Cox's paper. *J. R. Stat. Soc. Ser. B, Stat. Methodol.* **34** 216.

CAI, T. and CHENG, S. (2004). Semiparametric regression analysis for doubly censored data. *Biometrika* **91** 277–290. MR2081301 https://doi.org/10.1093/biomet/91.2.277

CANADIAN INSTITUTE FOR HEALTH INFORMATION National Ambulatory Care Reporting System (NACRS) metadata.

CANADIAN MENTAL HEALTH ASSOCIATION (2021). Fast facts about mental health and mental illness.

CHEN, A. A., HU, X. J. and ROSYCHUK, R. J. (2026). Supplement to "Stratified Regression Analysis of Zero-Truncated Recurrent Event Data." https://doi.org/10.1214/25-AOAS2130SUPP

COOK, R. J. and LAWLESS, J. F. (2007). *The Statistical Analysis of Recurrent Events*, 1st ed. *Statistics for Biology and Health*. Springer, New York. MR3822124

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

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

DOBLER, D., PAULY, M. and SCHEIKE, T. H. (2019). Confidence bands for multiplicative hazards models: Flexible resampling approaches. *Biometrics* **75** 906–916. MR4012096 https://doi.org/10.1111/biom.13059

EDDELBUETTEL, D., FRANCOIS, R., ALLAIRE, J., USHEY, K., KOU, Q., RUSSELL, N., UCAR, I., BATES, D. and CHAMBERS, J. (2025a). Rcpp: Seamless R and C++ integration. R package version 1.1.0.

EDDELBUETTEL, D., FRANCOIS, R., BATES, D., NI, B. and SANDERSON, C. (2025b). RcppArmadillo: 'Rcpp' integration for the 'Armadillo' templated linear algebra library. R package version 14.6.0-1.

HU, X. J. and LAWLESS, J. F. (1996). Estimation of rate and mean functions from truncated recurrent event data. *J. Amer. Statist. Assoc.* **91** 300–310. MR1394085 https://doi.org/10.2307/2291408

Key words and phrases. Conditional intensity function, doubly censored recurrent event data, health administrative data, partially known stratification, supplementary information.

HU, X. J. and ROSYCHUK, R. J. (2016). Marginal regression analysis of recurrent events with coarsened censoring times. *Biometrics* **72** 1113–1122. MR3591596 https://doi.org/10.1111/biom.12503

HUANG, C.-Y. and WANG, M.-C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *J. Amer. Statist. Assoc.* **99** 1153–1165. MR2109503 https://doi.org/10.1198/016214504000001033

LIN, D. Y., WEI, L. J. and YING, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80** 557–572. MR1248021 https://doi.org/10.1093/biomet/80.3.557

MASON, S., SPIWAK, R. and LOGSETTY, S. (2020). Population-Based Research Using Administrative Data to Evaluate Long-Term Outcomes in Burn Injury. *Handbook of Burns Volume* 1 85–92. Springer, Berlin.

NEWTON, A. S., ROSYCHUK, R. J., DONG, K., CURRAN, J., SLOMP, M. and MCGRATH, P. J. (2012). Emergency health care use and follow-up among sociodemographic groups of children who visit emergency departments for mental health crises. *CMAJ* **184** E665–E674.

NIRMALKANNA, K. and CIGSAR, C. (2024). Analysis of recurrent event processes with dynamic models for event counts. *Stat. Biosci.* 1–45.

PRENTICE, R. L., WILLIAMS, B. J. and PETERSON, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68** 373–379. MR0626396 https://doi.org/10.1093/biomet/68.2.373

SU, Y.-R. and WANG, J.-L. (2016). Semiparametric efficient estimation for shared-frailty models with doubly-censored clustered data. *Ann. Statist.* **44** 1298–1331. MR3485961 https://doi.org/10.1214/15-AOS1406

SUN, J. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*, 1st ed. *Statistics for Biology and Health*. Springer, New York. MR2287318

XIONG, Y., HU, J. and ROSYCHUK, R. (2026). Exploring differences between two decades of mental health related emergency department visits by youth via recurrent events analyses. *J. R. Stat. Soc. Ser. C, Appl. Stat.*.

ZHANG, C.-H. and LI, X. (1996). Linear regression with doubly censored data. *Ann. Statist.* **24** 2720–2743. MR1425976 https://doi.org/10.1214/aos/1032181177

ZHAO, W., PENG, L. and HANFELT, J. (2022). Semiparametric latent class analysis of recurrent event data. *J. R. Stat. Soc. Ser. B, Stat. Methodol.* **84** 1175–1197. MR4494157 https://doi.org/10.1111/rssb.12499

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 452–475
https://doi.org/10.1214/25-AOAS2131

A BLOCKWISE MIXED MEMBERSHIP MODEL FOR MULTIVARIATE LONGITUDINAL DATA: DISCOVERING CLINICAL HETEROGENEITY AND IDENTIFYING PARKINSON'S DISEASE SUBTYPES

BY KAI KANG[1,a] AND YUQI GU[2,b]

[1]*Department of Statistics, Sun Yat-sen University,* [a]*kangk5@mail.sysu.edu.cn*
[2]*Department of Statistics, Columbia University,* [b]*yuqi.gu@columbia.edu*

Current diagnosis and prognosis for Parkinson's disease (PD) face formidable challenges due to the heterogeneous nature of the disease course, including that: (i) the impairment severity varies hugely between patients, (ii) whether a symptom occur independently or co-occurs with related symptoms differs significantly, and (iii) repeated symptom measurements exhibit substantial temporal dependence. To tackle these challenges, we propose a novel blockwise mixed membership model (BM^3) to systematically unveil between-patient, between-symptom, and between-time clinical heterogeneity within PD. The key idea behind BM^3 is to partition multivariate longitudinal measurements into distinct blocks, enabling measurements within each block to share a common latent membership while allowing latent memberships to vary across blocks. Consequently, the heterogeneous PD-related measurements across time are divided into clinically homogeneous blocks consisting of correlated symptoms and consecutive time. From the analysis of Parkinson's Progression Markers Initiative data ($n = 1531$), we discover three typical disease profiles (stages), four symptom groups (i.e., autonomic function, tremor, left-side and right-side motor function), and two periods, advancing the comprehension of PD heterogeneity. Moreover, we identify several clinically meaningful PD subtypes by summarizing the blockwise latent memberships, paving the way for developing more precise and targeted therapies to benefit patients. Our findings are validated using external variables, successfully reproduced in validation datasets, and compared with existing methods. Theoretical results of model identifiability further ensure the reliability and reproducibility of latent structure discovery in PD.

REFERENCES

AIROLDI, E. M., BLEI, D. M., EROSHEVA, E. A. and FIENBERG, S. E. (2015). *Handbook of Mixed Membership Models and Their Applications*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR3381023

ALLMAN, E. S., MATIAS, C. and RHODES, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.* **37** 3099–3132. MR2549554 https://doi.org/10.1214/09-AOS689

BHATTACHARYA, A. and DUNSON, D. B. (2011). Sparse Bayesian infinite factor models. *Biometrika* **98** 291–306. MR2806429 https://doi.org/10.1093/biomet/asr013

CHEN, J. and KHALILI, A. (2009). Order selection in finite mixture models with a nonsmooth penalty. *J. Amer. Statist. Assoc.* **104** 187–196. MR2662302 https://doi.org/10.1198/jasa.2009.0103

EROSHEVA, E., FIENBERG, S. and LAFFERTY, J. (2004). Mixed-membership models of scientific publications. *Proc. Natl. Acad. Sci. USA* **101** 5220–5227.

EROSHEVA, E. A., FIENBERG, S. E. and JOUTARD, C. (2007). Describing disability through individual-level mixture models for multivariate binary data. *Ann. Appl. Stat.* **1** 502–537. MR2415745 https://doi.org/10.1214/07-AOAS126

FERESHTEHNEJAD, S.-M., ZEIGHAMI, Y., DAGHER, A. and POSTUMA, R. B. (2017). Clinical criteria for subtyping Parkinson's disease: Biomarkers and longitudinal progression. *Brain* **140** 1959–1976.

Key words and phrases. Latent variable model, Bayesian inference, disease subtypes, multivariate longitudinal data, identifiability, mixed membership model.

GELMAN, A., HWANG, J. and VEHTARI, A. (2014). Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24** 997–1016. MR3253850 https://doi.org/10.1007/s11222-013-9416-2

GENG, J., BHATTACHARYA, A. and PATI, D. (2019). Probabilistic community detection with unknown number of communities. *J. Amer. Statist. Assoc.* **114** 893–905. MR3963189 https://doi.org/10.1080/01621459.2018.1458618

GOETZ, C. G., TILLEY, B. C., SHAFTMAN, S. R., STEBBINS, G. T., FAHN, S., MARTINEZ-MARTIN, P., POEWE, W., SAMPAIO, C., STERN, M. B. et al. (2008). Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (mds-updrs): Scale presentation and clinimetric testing results. *Mov. Disord.* **23** 2129–2170.

GOODMAN, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61** 215–231. MR0370936 https://doi.org/10.1093/biomet/61.2.215

GREENLAND, J. C., WILLIAMS-GRAY, C. H. and BARKER, R. A. (2019). The clinical heterogeneity of Parkinson's disease and its therapeutic implications. *Eur. J. Neurosci.* **49** 328–338.

GU, Y., EROSHEVA, E. A., XU, G. and DUNSON, D. B. (2023). Dimension-grouped mixed membership models for multivariate categorical data. *J. Mach. Learn. Res.* **24** Paper No. [88], 49. MR4582510

HAGENAARS, J. A. and MCCUTCHEON, A. L. (2002). *Applied Latent Class Analysis* Cambridge Univ. Press, Cambridge. MR1927663 https://doi.org/10.1017/CBO9780511499531

HE, Y., SONG, X. and KANG, K. (2024). Joint mixed membership modeling of multivariate longitudinal and survival data for learning the individualized disease progression. *Ann. Appl. Stat.* **18** 1924–1946. MR4782472 https://doi.org/10.1214/23-aoas1864

HOEHN, M. M. and YAHR, M. D. (1967). Parkinsonism: Onset, progression, and mortality. *Neurology* **17** 427–427.

JACK, 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.

KANG, K. and GU, Y. (2026). Supplement to "A blockwise mixed membership model for multivariate longitudinal data: Discovering clinical heterogeneity and identifying Parkinson's disease subtypes." https://doi.org/10.1214/25-AOAS2131SUPPA, https://doi.org/10.1214/25-AOAS2131SUPPB

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

KRUSKAL, J. B. (1976). More factors than subjects, tests and treatments: An indeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika* **41** 281–293. MR0488592 https://doi.org/10.1007/BF02293554

KRUSKAL, J. B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.* **18** 95–138. MR0444690 https://doi.org/10.1016/0024-3795(77)90069-6

LAZARSFELD, P. F. and HENRY, N. W. (1968). *Latent Structure Analysis*. Houghton, Boston, MA.

LEWIS, S., FOLTYNIE, T., BLACKWELL, A. D., ROBBINS, T. W., OWEN, A. M. and BARKER, R. A. (2005). Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach. *J. Neurol. Neurosurg. Psychiatry* **76** 343–348.

MACQUEEN, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley, Calif.*, 1965/66), *Vol. I*: *Statistics* 281–297. Univ. California Press, Berkeley, CA. MR0214227

MANRIQUE-VALLIER, D. (2014). Longitudinal mixed membership trajectory models for disability survey data. *Ann. Appl. Stat.* **8** 2268–2291. MR3292497 https://doi.org/10.1214/14-AOAS769

MAREK, K., JENNINGS, D., LASCH, S., SIDEROWF, A., TANNER, C., SIMUNI, T., COFFEY, C., KIEBURTZ, K., FLAGG, E. et al. (2011). The Parkinson progression marker initiative (ppmi). *Prog. Neurobiol.* **95** 629–635.

MARRAS, C. (2015). Subtypes of Parkinson's disease: State of the field and future directions. *Curr. Opin. Neurol.* **28** 382–386.

MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models*. *Wiley Series in Probability and Statistics*: *Applied Probability and Statistics*. Wiley-Interscience, New York. MR1789474 https://doi.org/10.1002/0471721182

MILLER, J. W. and HARRISON, M. T. (2018). Mixture models with a prior on the number of components. *J. Amer. Statist. Assoc.* **113** 340–356. MR3803469 https://doi.org/10.1080/01621459.2016.1255636

PIGOTT, K., RICK, J., XIE, S. X., HURTIG, H., CHEN-PLOTKIN, A., DUDA, J. E., MORLEY, J. F., CHAHINE, L. M., DAHODWALA, N. et al. (2015). Longitudinal study of normal cognition in Parkinson disease. *Neurology* **85** 1276–1282.

RAFTERY, A. E., NEWTON, M. A., SATAGOPAN, J. M. and KRIVITSKY, P. N. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In *Bayesian Statistics*, *Vol.* 8. *Oxford Sci. Publ.* 371–416. Oxford Univ. Press, Oxford. MR2433201

RAND, W. M. (1971). Objective criteria for the evaluation of clustering methods. *J. Amer. Statist. Assoc.* **66** 846–850.

SEVERSON, K. A., CHAHINE, L. M., SMOLENSKY, L., NG, K., HU, J. and GHOSH, S. (2020). Personalized input-output hidden Markov models for disease progression modeling. In *Machine Learning for Healthcare Conference* 309–330. PMLR.

THENGANATT, M. A. and JANKOVIC, J. (2014). Parkinson disease subtypes. *JAMA Neurol*. **71** 499–504.

TOLOSA, E., GAIG, C., SANTAMARÍA, J. and COMPTA, Y. (2009). Diagnosis and the premotor phase of Parkinson disease. *Neurology* **72** S12–S20.

WANG, Q. and WANG, Y. (2024). Multilayer exponential family factor models for integrative analysis and learning disease progression. *Biostatistics* **25** 203–219. MR4678542 https://doi.org/10.1093/biostatistics/kxac042

WANG, Y. S., MATSUEDA, R. L. and EROSHEVA, E. A. (2017). A variational EM method for mixed membership models with multivariate rank data: An analysis of public policy preferences. *Ann. Appl. Stat.* **11** 1452–1480. MR3709566 https://doi.org/10.1214/17-AOAS1034

WATANABE, S. and OPPER, M. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11** 3571–3594. MR2756194

WOODBURY, M. A., CLIVE, J. and GARSON, A. JR (1978). Mathematical typology: A grade of membership technique for obtaining disease definition. *Comput. Biomed. Res.* **11** 277–298.

ZHOU, M. (2018). Nonparametric Bayesian negative binomial factor analysis. *Bayesian Anal.* **13** 1065–1093. MR3855363 https://doi.org/10.1214/17-BA1070

ZHOU, M., HANNAH, L., DUNSON, D. and CARIN, L. (2012). Beta-negative binomial process and Poisson factor analysis. In *Proceedings of the* 15*th International Conference on Artificial Intelligence and Statistics* (*AISTATS*) 1462–1471. PMLR.

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 476–495
https://doi.org/10.1214/26-AOAS2139

MULTILEVEL FUNCTIONAL DISTRIBUTIONAL MODELS WITH APPLICATIONS TO CONTINUOUS GLUCOSE MONITORING IN DIABETES CLINICAL TRIALS

BY MARCOS MATABUENA[1,a] AND CIPRIAN M. CRAINICEANU[2,b]

[1]*Mohamed bin Zayed University of Artificial Intelligence*, [a]*Marcos.Matabuena@mbzuai.ac.ae*
[2]*Department of Biostatistics, Johns Hopkins University*, [b]*ccraini1@jhu.edu*

Continuous glucose monitoring (CGM) is a minimally invasive technology that measures blood glucose every few minutes for weeks or months at a time. CGM data are often collected in the free-living environment and is strongly related to sleep, physical activity, and meal intake. As the timing of these activities varies substantially within- and between-individuals, it is difficult to model CGM trajectories as a function of time of day. Therefore, in practice, CGM trajectories are often reduced to one or two scalar summaries of the thousands of measurements collected for a study participant. To alleviate the potential loss of information, the cumulative distribution function (cdf) of the CGM time series was proposed as an alternative. Here we address the problem of conducting inference on cdfs in clinical trials with long follow-up and frequent measurements. Our approach provides three major innovations: (1) modeling the entire cdf and preserving its monotonicity, (2) accounting for the cdfs correlation (because they are measured on the same individual), continuity (results are robust to the choice of the probability grid), and differential error (e.g., medians have lower variability than 0.99 quantiles), and (3) preserving the familywise error when the observed data are longitudinal samples of cdfs. We focus on modeling data collected by The Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Group in a large clinical trial that collected CGM data every few minutes for 26 weeks. Our basic observation unit is the distribution of CGM observations in a four–week interval. The resulting data structure is multilevel (because each individual has multiple months of data) and distributional (because the data for each four-week interval is represented as a cdf). The scientific goals are to: (1) identify and quantify the effects of factors that affect glycaemic control in type 1 diabetes patients (T1D) and (2) identify and characterize the patients who respond to treatment.

REFERENCES

AJJAN, R., SLATTERY, D. and WRIGHT, E. (2019). Continuous glucose monitoring: A brief review for primary care practitioners. *Adv. Ther.* **36** 579–596. https://doi.org/10.1007/s12325-019-0870-x

AJJAN, R. A. (2017). How can we realize the clinical benefits of continuous glucose monitoring? *Diabetes Technol. Ther.* **19** S27–S36. https://doi.org/10.1089/dia.2017.0021

BAILEY, T., BODE, B. W., CHRISTIANSEN, M. P., KLAFF, L. J. and ALVA, S. (2015). The performance and usability of a factory-calibrated flash glucose monitoring system. *Diabetes Technol. Ther.* **17** 787–794. https://doi.org/10.1089/dia.2014.0378

BATTELINO, T., ALEXANDER, C. M., AMIEL, S. A., ARREAZA-RUBIN, G., BECK, R. W., BERGENSTAL, R. M., BUCKINGHAM, B. A., CARROLL, J., CERIELLO, A. et al. (2022). Continuous glucose monitoring and metrics for clinical trials: An international consensus statement. *Lancet Diabetes Endocrinol.*

BECK, R. W., BERGENSTAL, R. M., RIDDLESWORTH, T. D., KOLLMAN, C., LI, Z., BROWN, A. S. and CLOSE, K. L. (2018). Validation of time in range as an outcome measure for diabetes clinical trials. *Diabetes Care* **42** 400–405. https://doi.org/10.2337/dc18-1444

Key words and phrases. Digital health, distributional data analysis, continuous glucose monitoring data, clinical trials.

Beck, R. W., Calhoun, P. and Kollman, C. (2012). Use of continuous glucose monitoring as an outcome measure in clinical trials. *Diabetes Technol. Ther.* **14** 877–882. https://doi.org/10.1089/dia.2012.0079

Ben-Yacov, O., Godneva, A., Rein, M., Shilo, S., Kolobkov, D., Koren, N., Cohen Dolev, N., Travinsky Shmul, T., Wolf, B. C. et al. (2021). Personalized postprandial glucose response–targeting diet versus Mediterranean diet for glycemic control in prediabetes. *Diabetes Care* **44** 1980–1991.

Brito, P. and Dias, S. (2022). *Analysis of Distributional Data*. CRC Press, Boca Raton.

Burge, M. R., Mitchell, S., Sawyer, A. and Schade, D. S. (2008). Continuous glucose monitoring: The future of diabetes management. *Diabetes Spectr.* **21** 112–119. https://doi.org/10.2337/diaspect.21.2.112

Crainiceanu, C. M., Goldsmith, J., Leroux, A. and Cui, E. (2024). *Functional Data Analysis with R*. CRC Press, Boca Raton.

Crainiceanu, C. M., Staicu, A.-M., Ray, S. and Punjabi, N. (2012). Bootstrap-based inference on the difference in the means of two correlated functional processes. *Stat. Med.* **31** 3223–3240. MR2993623 https://doi.org/10.1002/sim.5439

Cui, E., Leroux, A., Smirnova, E. and Crainiceanu, C. M. (2022). Fast univariate inference for longitudinal functional models. *J. Comput. Graph. Statist.* **31** 219–230. MR4387222 https://doi.org/10.1080/10618600.2021.1950006

Cui, E., Li, R., Crainiceanu, C. M. and Xiao, L. (2023a). Fast multilevel functional principal component analysis. *J. Comput. Graph. Statist.* **32** 366–377. MR4592916 https://doi.org/10.1080/10618600.2022.2115500

Cui, E. H., Goldfine, A., Quinlan, M., James, D. and Sverdlov, O. (2023b). Investigating the value of glucodensity analysis of continuous glucose monitoring data in type 1 diabetes: An exploratory analysis. *Front. Clin. Diabetes Healthc.* **4** 1244613.

Gaynanova, I., Punjabi, N. and Crainiceanu, C. (2022). Modeling continuous glucose monitoring (CGM) data during sleep. *Biostatistics* **23** 223–239. MR4366045 https://doi.org/10.1093/biostatistics/kxaa023

Gertheiss, J., Goldsmith, J., Crainiceanu, C. and Greven, S. (2013). Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics* **14** 447–461. https://doi.org/10.1093/biostatistics/kxs051

Ghosal, A., Matabuena, M., Meiring, W. and Petersen, A. (2023a). Predicting distributional profiles of physical activity in the NHANES database using a Partially Linear Single-Index Fréchet Regression model. arXiv preprint. Available at arXiv:2302.07692.

Ghosal, R., Ghosh, S., Urbanek, J., Schrack, J. A. and Zipunnikov, V. (2023b). Shape-constrained estimation in functional regression with Bernstein polynomials. *Comput. Statist. Data Anal.* **178** Paper No. 107614, 20. MR4483316 https://doi.org/10.1016/j.csda.2022.107614

Ghosal, R. and Matabuena, M. (2023). Multivariate scalar on multidimensional distribution regression. arXiv preprint. Available at arXiv:2310.10494.

Ghosal, R., Varma, V. R., Volfson, D., Hillel, I., Urbanek, J., Hausdorff, J. M., Watts, A. and Zipunnikov, V. (2021a). Distributional data analysis via quantile functions and its application to modelling digital biomarkers of gait in Alzheimer's Disease.

Ghosal, R., Varma, V. R., Volfson, D., Urbanek, J., Hausdorff, J. M., Watts, A. and Zipunnikov, V. (2021b). Scalar on time-by-distribution regression and its application for modelling associations between daily-living physical activity and cognitive functions in Alzheimer's Disease.

Goldsmith, J., Crainiceanu, C. M., Caffo, B. and Reich, D. (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **61** 453–469. MR2914521 https://doi.org/10.1111/j.1467-9876.2011.01031.x

Greven, S., Crainiceanu, C., Caffo, B. and Reich, D. (2010). Longitudinal functional principal component analysis. *Electron. J. Stat.* **4** 1022–1054. MR2727452 https://doi.org/10.1214/10-EJS575

Jeong, S., Morris, J. and Yang, H. (2024). Warsserstein Tests for Equality of Several Groups for Distributional Data. https://doi.org/10.21203/rs.3.rs-4536450/v1

Koffman, L., Crainiceanu, C. and Leroux, A. (2024). Walking fingerprinting. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **73** 1221–1241. MR4896621 https://doi.org/10.1093/jrsssc/qlae033

Leshem, A., Segal, E. and Elinav, E. (2020). The gut microbiome and individual-specific responses to diet. *mSystems* **5** e00665–20.

Liao, Y., Thompson, C., Peterson, S., Mandrola, J. and Beg, M. S. (2019). The future of wearable technologies and remote monitoring in health care. *Amer. Soc. Clin. Oncol. Educ. Book* **39** 115–121.

Martens, T. W., Bergenstal, R. M., Pearson, T., Carlson, A. L., Scheiner, G., Carlos, C., Liao, B., Syring, K. and Pollom, R. D. (2021). Making sense of glucose metrics in diabetes: Linkage between postprandial glucose (PPG), time in range (TIR) & hemoglobin A1C (A1C). *Postgrad. Med.* **133** 253–264.

Matabuena, M., Félix, P., Ditzhaus, M., Vidal, J. and Gude, F. (2023a). Hypothesis testing for matched pairs with missing data by maximum mean discrepancy: An application to continuous glucose monitoring. *Amer. Statist.* **77** 357–369. MR4661582 https://doi.org/10.1080/00031305.2023.2200512

MATABUENA, M., FELIX, P., GARCIA-MEIXIDE, C. and GUDE, F. (2022a). Kernel machine learning methods to handle missing responses with complex predictors. Application in modelling five-year glucose changes using distributional representations. *Comput. Methods Programs Biomed.* **221** 106905.

MATABUENA, M., FÉLIX, P., HAMMOURI, Z. A. A., MOTA, J. and DEL POZO CRUZ, B. (2022b). Physical activity phenotypes and mortality in older adults: A novel distributional data analysis of accelerometry in the NHANES. *Aging Clin. Exp. Res.* https://doi.org/10.1007/s40520-022-02260-3

MATABUENA, M., GHOSAL, R., AGUILAR, J. E., KESHET, A., WAGNER, R., FERNÁNDEZ MERINO, C., SÁNCHEZ CASTRO, J., ZIPUNNIKOV, V., ONNELA, J.-P. et al. (2025). Glucodensity functional profiles outperform traditional continuous glucose monitoring metrics. *Sci. Rep.* **15** 33662. https://doi.org/10.1038/s41598-025-18119-2

MATABUENA, M., KARAS, M., RIAZATI, S., CAPLAN, N. and HAYES, P. R. (2023b). Estimating knee movement patterns of recreational runners across training sessions using multilevel functional regression models. *Amer. Statist.* **77** 169–181. MR4579291 https://doi.org/10.1080/00031305.2022.2105950

MATABUENA, M. and PETERSEN, A. (2023). Distributional data analysis of accelerometer data from the NHANES database using nonparametric survey regression models. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **72** 294–313. MR4719277 https://doi.org/10.1093/jrsssc/qlad007

MATABUENA, M., PETERSEN, A., VIDAL, J. C. and GUDE, F. (2021). Glucodensities: A new representation of glucose profiles using distributional data analysis. *Stat. Methods Med. Res.* **30** 1445–1464. MR4269959 https://doi.org/10.1177/0962280221998064

MATABUENA, M., SARTINI, J. and GUDE, F. (2024). Multilevel functional data analysis modeling of human glucose response to meal intake.

NØRGAARD, K., RANJAN, A. G., LAUGESEN, C., TIDEMAND, K. G., GREEN, A., SELMER, C., SVENSSON, J., ANDERSEN, H. U., VISTISEN, D. et al. (2023). Glucose monitoring metrics in individuals with type 1 diabetes using different treatment modalities: A real-world observational study. *Diabetes Care* dc231137. https://doi.org/10.2337/dc23-1137

PETERSEN, A., LIU, X. and DIVANI, A. A. (2021). Wasserstein F-tests and confidence bands for the Fréchet regression of density response curves. *Ann. Statist.* **49** 590–611. MR4206692 https://doi.org/10.1214/20-AOS1971

REISS, P. T., HUANG, L. and MENNES, M. (2010). Fast function-on-scalar regression with penalized basis expansions. *Int. J. Biostat.* **6** Art. 28, 30. MR2683940 https://doi.org/10.2202/1557-4679.1246

RODBARD, D. (2016). Continuous glucose monitoring: A review of successes, challenges, and opportunities. *Diabetes Technol. Ther.* **18** S2-3–S2-13. https://doi.org/10.1089/dia.2015.0417

RODBARD, D. (2017). Continuous glucose monitoring: A review of recent studies demonstrating improved glycemic outcomes. *Diabetes Technol. Ther.* **19** S-25–S-37. https://doi.org/10.1089/dia.2017.0035

SCHEIPL, F., STAICU, A.-M. and GREVEN, S. (2015). Functional additive mixed models. *J. Comput. Graph. Statist.* **24** 477–501. MR3357391 https://doi.org/10.1080/10618600.2014.901914

SCHNELL, O., BARNARD, K., BERGENSTAL, R., BOSI, E., GARG, S., GUERCI, B., HAAK, T., HIRSCH, I. B., JI, L. et al. (2017). Role of continuous glucose monitoring in clinical trials: Recommendations on reporting. *Diabetes Technol. Ther.* **19** 391–399. https://doi.org/10.1089/dia.2017.0054

SERGAZINOV, R., LEROUX, A., CUI, E., CRAINICEANU, C., AURORA, R. N., PUNJABI, N. M. and GAYNANOVA, I. (2022). A case study of glucose levels during sleep using fast function on scalar regression inference. arXiv preprint. Available at arXiv:2205.08439.

SERGAZINOV, R., LEROUX, A., CUI, E., CRAINICEANU, C., AURORA, R. N., PUNJABI, N. M. and GAYNANOVA, I. (2023). A case study of glucose levels during sleep using multilevel fast function on scalar regression inference. *Biometrics* **79** 3873–3882. MR4680766 https://doi.org/10.1111/biom.13878

SHOU, H., ZIPUNNIKOV, V., CRAINICEANU, C. M. and GREVEN, S. (2015). Structured functional principal component analysis. *Biometrics* **71** 247–257. MR3335369 https://doi.org/10.1111/biom.12236

STAICU, A.-M., CRAINICEANU, C. M. and CARROLL, R. J. (2010). Fast methods for spatially correlated multilevel functional data. *Biostatistics* **11** 177–194. https://doi.org/10.1093/biostatistics/kxp058

STAICU, A.-M., ISLAM, M. N., DUMITRU, R. and VAN HEUGTEN, E. (2020). Longitudinal dynamic functional regression. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **69** 25–46. MR4052797 https://doi.org/10.1111/rssc.12376

TAMBORLANE, W. V., BECK, R. W., BODE, B. W. et al. (2008). Continuous glucose monitoring and intensive treatment of type 1 diabetes. *N. Engl. J. Med.* **359** 1464–1476.

TANG, B., ZHAO, Y., VENKATARAMAN, A., TSAPKINI, K., LINDQUIST, M. A., PEKAR, J. and CAFFO, B. (2020). Differences in functional connectivity distribution after transcranial direct-current stimulation: A connectivity density point of view. bioRxiv.

THE JUVENILE DIABETES RESEARCH FOUNDATION CONTINUOUS GLUCOSE MONITORING STUDY GROUP (2009). The effect of continuous glucose monitoring in well-controlled type 1 diabetes. *Diabetes Care* **32** 1378–1383.

WOOD, A., O'NEAL, D., FURLER, J. and EKINCI, E. I. (2018). Continuous glucose monitoring: A review of the evidence, opportunities for future use and ongoing challenges. *Intern. Med. J.* **48** 499–508. https://doi.org/10.1111/imj.13770

YANG, H., BALADANDAYUTHAPANI, V., RAO, A. U. K. and MORRIS, J. S. (2020). Quantile function on scalar regression analysis for distributional data. *J. Amer. Statist. Assoc.* **115** 90–106. MR4078447 https://doi.org/10.1080/01621459.2019.1609969

YOO, J. H. and KIM, J. H. (2020). Time in range from continuous glucose monitoring: A novel metric for glycemic control. *Diabetes Metab. J.* **44** 828–839.

ZIPUNNIKOV, V., GREVEN, S., SHOU, H., CAFFO, B. S., REICH, D. S. and CRAINICEANU, C. M. (2014). Longitudinal high-dimensional principal components analysis with application to diffusion tensor imaging of multiple sclerosis. *Ann. Appl. Stat.* **8** 2175–2202. MR3292493 https://doi.org/10.1214/14-AOAS748

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 496–510
https://doi.org/10.1214/26-AOAS2144
© Institute of Mathematical Statistics, 2026

REGRESSION ANALYSIS OF CASE *K* INTERVAL-CENSORED FAILURE TIME DATA WITH RANDOM CHANGE POINT AND INFORMATIVE CENSORING

BY MINGYUE DU[1,a], YICHEN LOU[2,b] AND JIANGUO SUN[3,c]

[1]*School of Mathematics, Jilin University,* [a]*mingydu@jlu.edu.cn*
[2]*School of Physical and Mathematical Sciences, Nanyang Technological University,* [b]*louyichen19@outlook.com*
[3]*Department of Statistics and Data Science, Southern University of Science and Technology,* [c]*suncolumbia@163.com*

This paper discusses regression analysis of interval-censored failure time data, which often occur in many areas and for which a great deal of literature has been established. In addition, many authors have investigated the analysis of failure time data with either change points or informative censoring, and as interval censoring, both can also separately occur in many situations such as clinical medicine and precision medicine. However, it does not seem to exist an established approach that can deal with the situation where all of the three issues occur together. To address this, we propose a sieve maximum likelihood estimation approach for regression analysis of case K interval-censored data in the presence of both a random change point and informative censoring. For the implementation of the proposed method, an EM algorithm is developed and also the asymptotic properties of the resulting estimators of regression parameters are established. Furthermore, a simulation study is conducted to assess the finite sample performance of the proposed method and suggests that it works well for practical situations. The method is applied to a set of breast cancer data that motivated this study.

REFERENCES

ADAMI, H.-O., MALKER, B., HOLMBERG, L., PERSSON, I. and STONE, B. (1986). The relation between survival and age at diagnosis in breast cancer. *N. Engl. J. Med.* **315** 559–563.

CHEN, M.-H., TONG, X. and SUN, J. (2009). A frailty model approach for regression analysis of multivariate current status data. *Stat. Med.* **28** 3424–3436. MR2744372 https://doi.org/10.1002/sim.3715

CHEN, X., PING, Y. and SUN, J. (2024). Efficient estimation of Cox model with random change point. *Stat. Med.* **43** 1213–1226. MR4707615 https://doi.org/10.1002/sim.9987

CHEN, Y., FENG, Y. and SUN, J. (2015). Regression analysis of multivariate current status data with auxiliary covariates under the additive hazards model. *Comput. Statist. Data Anal.* **87** 34–45. MR3319805 https://doi.org/10.1016/j.csda.2015.01.005

CHLEBOWSKI, R. T., MANSON, J. E., ANDERSON, G. L., CAULEY, J. A., ARAGAKI, A. K., STEFANICK, M. L., LANE, D. S., JOHNSON, K. C., WACTAWSKI-WENDE, J. et al. (2013). Estrogen plus progestin and breast cancer incidence and mortality in the women's health initiative observational study. *J. Natl. Cancer Inst.* **105** 526–535.

COLDITZ, G. A., ROSNER, B. A., CHEN, W. Y., HOLMES, M. D. and HANKINSON, S. E. (2004). Risk factors for breast cancer according to estrogen and progesterone receptor status. *J. Natl. Cancer Inst.* **96** 218–228.

DALL, G. V. and BRITT, K. L. (2017). Estrogen effects on the mammary gland in early and late life and breast cancer risk. *Front. Oncol.* **7** 110.

DU, M., LOU, Y. and SUN, J. (2026). Supplement to "Regression analysis of case K interval-censored failure time data with random change point and informative censoring." https://doi.org/10.1214/26-AOAS2144SUPP

EFRON, B. (1992). Bootstrap methods: Another look at the jackknife. In *Breakthroughs in Statistics*: *Methodology and Distribution* 569–593. Springer, Berlin. https://doi.org/10.1007/978-1-4612-4380-9_41

FARRINGTON, C. and GAY, N. (1999). Interval-censored survival data with informative examination times: Parametric models and approximate inference. *Stat. Med.* **18** 1235–1248.

Key words and phrases. Cox model, EM algorithm, informative interval censoring, random change point, sieve maximum likelihood estimation.

FINKELSTEIN, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42** 845–854. MR0872963 https://doi.org/10.2307/2530698

FINKELSTEIN, D. M., GOGGINS, W. B. and SCHOENFELD, D. A. (2002). Analysis of failure time data with dependent interval censoring. *Biometrics* **58** 298–304. MR1908169 https://doi.org/10.1111/j.0006-341X.2002.00298.x

HEER, E., HARPER, A., ESCANDOR, N., SUNG, H., MCCORMACK, V. and FIDLER-BENAOUDIA, M. M. (2020). Global burden and trends in premenopausal and postmenopausal breast cancer: A population-based study. *Lancet Glob. Health* **8** e1027–e1037.

HU, T., ZHOU, Q. and SUN, J. (2017). Regression analysis of bivariate current status data under the proportional hazards model. *Canad. J. Statist.* **45** 410–424. MR3729978 https://doi.org/10.1002/cjs.11344

HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24** 540–568. MR1394975 https://doi.org/10.1214/aos/1032894452

JENSEN, U. and LÜTKEBOHMERT, C. (2008). A Cox-type regression model with change-points in the covariates. *Lifetime Data Anal.* **14** 267–285. MR2426673 https://doi.org/10.1007/s10985-008-9083-3

KALBFLEISCH, J. D. and PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR1924807 https://doi.org/10.1002/9781118032985

LAM, K. F., XU, J. and XUE, H. (2018). Estimation of age effect with change-points on survival of cancer patients. *Stat. Med.* **37** 1732–1743. MR3787984 https://doi.org/10.1002/sim.7618

LEE, C. Y. and LAM, K. F. (2020). Survival analysis with change-points in covariate effects. *Stat. Methods Med. Res.* **29** 3235–3248. MR4156851 https://doi.org/10.1177/0962280220922258

LI, S., HU, T., WANG, P. and SUN, J. (2017). Regression analysis of current status data in the presence of dependent censoring with applications to tumorigenicity experiments. *Comput. Statist. Data Anal.* **110** 75–86. MR3612609 https://doi.org/10.1016/j.csda.2016.12.011

LIU, Y., HU, T. and SUN, J. (2017). Regression analysis of current status data in the presence of a cured subgroup and dependent censoring. *Lifetime Data Anal.* **23** 626–650. MR3705828 https://doi.org/10.1007/s10985-016-9382-z

MA, L., HU, T. and SUN, J. (2015). Sieve maximum likelihood regression analysis of dependent current status data. *Biometrika* **102** 731–738. MR3394289 https://doi.org/10.1093/biomet/asv020

SAMAAN, S. A. and CRAWFORD, M. H. (1995). Estrogen and cardiovascular function after menopause. *J. Amer. Coll. Cardiol.* **26** 1403–1410.

SIEGEL, R. L., GIAQUINTO, A. N. and JEMAL, A. (2024). Cancer statistics. *CA Cancer J. Clin.* **74** 12–49. https://doi.org/10.3322/caac.21820

SUN, J. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*. *Statistics for Biology and Health*. Springer, New York. MR2287318

SUN, J. and CHEN, D.-G. (2022). *Emerging Topics in Modeling Interval-Censored Survival Data*. Springer, Cham, Switzerland.

TENG, J. (2019). SEER breast cancer data. IEEE DataPort. https://doi.org/10.21227/a9qy-ph35

WANG, P., ZHAO, H. and SUN, J. (2016). Regression analysis of case K interval-censored failure time data in the presence of informative censoring. *Biometrics* **72** 1103–1112. MR3591595 https://doi.org/10.1111/biom.12527

YIN LEE, C. and WONG, K. Y. (2023). Survival analysis with a random change-point. *Stat. Methods Med. Res.* **32** 2083–2095. MR4671000 https://doi.org/10.1177/09622802231192946

YIP, C.-H. and RHODES, A. (2014). Estrogen and progesterone receptors in breast cancer. *Future Oncol.* **10** 2293–2301.

ZHANG, Z., SUN, J. and SUN, L. (2005). Statistical analysis of current status data with informative observation times. *Stat. Med.* **24** 1399–1407. MR2134566 https://doi.org/10.1002/sim.2001

ZHANG, Z., SUN, L., SUN, J. and FINKELSTEIN, D. M. (2007). Regression analysis of failure time data with informative interval censoring. *Stat. Med.* **26** 2533–2546. MR2361363 https://doi.org/10.1002/sim.2721

ZHAO, S., HU, T., MA, L., WANG, P. and SUN, J. (2015). Regression analysis of interval-censored failure time data with the additive hazards model in the presence of informative censoring. *Stat. Interface* **8** 367–377. MR3341334 https://doi.org/10.4310/SII.2015.v8.n3.a10

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 511–535
https://doi.org/10.1214/25-AOAS2054
© Institute of Mathematical Statistics, 2026

ASSESSING INFLUENTIAL OBSERVATIONS IN PAIN PREDICTION USING FMRI DATA

BY DONGLIANG ZHANG[1,a], MASOUD ASGHARIAN[2,c] AND MARTIN A. LINDQUIST[1,b]

[1]*Department of Biostatistics, Johns Hopkins University,* [a]*dzhang69@jhu.edu,* [b]*mlindqui@jhsph.edu*
[2]*Department of Mathematics and Statistics, McGill University,* [c]*masoud.asgharian2@mcgill.ca*

Neuroimaging data allows researchers to model the relationship between multivariate patterns of brain activity and outcomes related to mental states and behaviors. However, the existence of outlying participants can potentially undermine the generalizability of these models and jeopardize the validity of downstream statistical analysis. To date, the ability to detect and account for participants unduly influencing various model selection approaches have been sorely lacking. Motivated by a task-based functional magnetic resonance imaging (fMRI) study of thermal pain, we propose and establish the asymptotic distribution for a diagnostic measure applicable to a number of different model selectors. A high-dimensional clustering procedure is further combined with this measure to detect multiple influential observations. In a series of simulations, our proposed method demonstrates clear advantages over existing methods in terms of improved detection performance, leading to enhanced predictive and variable selection outcomes. Application of our method to data from the thermal pain study illustrates the influence of outlying participants, in particular with regards to differences in activation between low and intense pain conditions. This allows for the selection of an interpretable model with high prediction power after removal of the detected observations. Though inspired by the fMRI-based thermal pain study, our methods are broadly applicable to other high-dimensional data types.

REFERENCES

ARTHUR, D. and VASSILVITSKII, S. (2007). K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 1027–1035. ACM, New York. MR2485254

ASHBURNER, J. and FRISTON, K. J. (2005). Unified segmentation. *NeuroImage* **26** 839–851.

BELSLEY, D. A., KUH, E. and WELSCH, R. E. (1980). *Regression Diagnostics*: *Identifying Influential Data and Sources of Collinearity*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0576408

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*, *Methodol*. **57** 289–300. MR1325392

BEYER, K., GOLDSTEIN, J., RAMAKRISHNAN, R. and SHAFT, U. (1999). When is "nearest neighbor" meaningful? In *Database Theory-ICDT* 1999 217–235.

BREHENY, P. and HUANG, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat.* **5** 232–253. MR2810396 https://doi.org/10.1214/10-AOAS388

COMMENGES, D. (2003). Transformations which preserve exchangeability and application to permutation tests. *J. Nonparametr. Stat.* **15** 171–185. MR1981459 https://doi.org/10.1080/1048525031000089310

COOK, R. D. (1977). Detection of influential observation in linear regression. *Technometrics* **19** 15–18. MR0436478 https://doi.org/10.2307/1268249

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Ph. D. qualifying paper, Dept. Statistics, Harvard Univ.

FAIRHURST, M., BERNA, K. F. C. and TRACEY, I. (2012). An fMRI study exploring the overlap and differences between neural representations of physical and recalled pain. *PLoS ONE* **7** 1–10.

FAN, J., HAN, F. and LIU, H. (2014). Challenges of big data analysis. *Nat. Sci. Rev.* **1** 293–314.

Key words and phrases. Regression diagnostics, group deletion, cluster analysis, exchangeability, fMRI.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

Freedman, D. A. (1983). A note on screening regression equations. *Amer. Statist.* **37** 152–155. MR0702208 https://doi.org/10.2307/2685877

Galambos, J. (1973). A general Poisson limit theorem of probability theory. *Duke Math. J.* **40** 581–586. MR0322930

Hadi, A. S. and Simonoff, J. S. (1993). Procedures for the identification of multiple outliers in linear models. *J. Amer. Statist. Assoc.* **88** 1264–1272. MR1245359

Hahn, M. G. and Zhang, G. (1998). Distinctions between the regular and empirical central limit theorems for exchangeable random variables. In *High Dimensional Probability* (*Oberwolfach*, 1996). *Progress in Probability* **43** 111–143. Birkhäuser, Basel. MR1652324

Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896. MR0301858 https://doi.org/10.1214/aoms/1177693054

Haynes, J.-D. (2015). A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives. *Neuron* **87** 257–270.

Hinton, G. E. and Roweis, S. (2002). Stochastic neighbor embedding. In *Adv. Neural Inf. Process. Syst* **15**.

Klass, M. and Teicher, H. (1987). The central limit theorem for exchangeable random variables without moments. *Ann. Probab.* **15** 138–153. MR0877594

Léger, C. and Altman, N. (1993). Assessing influence in variable selection problems. *J. Amer. Statist. Assoc.* **88** 547–556. MR1224380

Lindquist, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. MR2530545 https://doi.org/10.1214/09-STS282

Lindquist, M. A., Krishnan, A., López-Solà, M., Jepma, M., Woo, C.-W., Koban, L., Roy, M., Atlas, L. Y., Schmidt, L. et al. (2017). Group-regularized individual prediction: Theory and application to pain. *NeuroImage* **145** 274–287.

Mejia, A. F., Nebel, M. B., Eloyan, A., Caffo, B. and Lindquist, M. A. (2017). PCA leverage: Outlier detection for high-dimensional functional magnetic resonance imaging data. *Biostatistics* **18** 521–536. MR3799592 https://doi.org/10.1093/biostatistics/kxw050

Ng, A., Jordan, M. and Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In *Adv. Neural Inf. Process. Syst* **14**.

Ombao, H., Lindquist, M., Thompson, W. and Aston, J. (2016). *Handbook of Neuroimaging Data Analysis*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR3618824

Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G. and Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neurosci. Biobehav. Rev.* **36** 1140–1152.

Pollard, D. (1981). Strong consistency of k-means clustering. *Ann. Statist.* **9** 135–140. MR0600539

Rajaratnam, B., Roberts, S., Sparks, D. and Yu, H. (2019). Influence diagnostics for high-dimensional lLasso regression. *J. Comput. Graph. Statist.* **28** 877–890. MR4045855 https://doi.org/10.1080/10618600.2019.1598869

Sun, T. and Zhang, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. MR2999166 https://doi.org/10.1093/biomet/ass043

Sun, W., Wang, J. and Fang, Y. (2012). Regularized k-means clustering of high-dimensional data and its asymptotic consistency. *Electron. J. Stat.* **6** 148–167. MR2879675 https://doi.org/10.1214/12-EJS668

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

Wager, T. D., Atlas, L. Y., Lindquist, M., Roy, M., Woo, C.-W. and Kross, E. (2013a). An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* **368** 1388–1397.

Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C.-W. and Kross, E. (2013b). An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* **368** 1388–1397.

Wang, G., Datta, A. and Lindquist, M. A. (2022). Bayesian functional registration of fMRI activation maps. *Ann. Appl. Stat.* **16** 1676–1699. MR4455896 https://doi.org/10.1214/21-aoas1562

Wang, G., Datta, A. and Lindquist, M. A. (2024). Improved fMRI-based pain prediction using Bayesian group-wise functional registration. *Biostatistics* **25** 885–903. MR4772536 https://doi.org/10.1093/biostatistics/kxad026

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

Zhang, D., Asgharian, M. and Lindquist, M. A. (2026). Supplement to “Assessing Influential Observations in Pain Prediction using fMRI Data.” https://doi.org/10.1214/25-AOAS2054SUPPA, https://doi.org/10.1214/25-AOAS2054SUPPB, https://doi.org/10.1214/25-AOAS2054SUPPC, https://doi.org/10.1214/25-AOAS2054SUPPD

ZHAO, J., LENG, C., LI, L. and WANG, H. (2013). High-dimensional influence measure. *Ann. Statist.* **41** 2639–2667. MR3161440 https://doi.org/10.1214/13-AOS1165

ZHAO, J., LIU, C., NIU, L. and LENG, C. (2019). Multiple influential point detection in high dimensional regression spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 385–408. MR3928147 https://doi.org/10.1111/rssb.12311

ZHAO, J., ZHANG, Y. and NIU, L. (2015). Detecting multiple influential observations in high dimensional linear regression. In *Advanced Intelligent Computing Theories and Applications* 55–64. Springer, Berlin.

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. MR2137327 https://doi.org/10.1111/j.1467-9868.2005.00503.x

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 536–559
https://doi.org/10.1214/25-AOAS2108
© Institute of Mathematical Statistics, 2026

BAYESIAN IMAGE-ON-IMAGE REGRESSION VIA DEEP KERNEL LEARNING BASED GAUSSIAN PROCESSES

BY GUOXUAN MA[1,a], BANGYAO ZHAO[1,b], HASAN ABU-AMARA[2,d] AND JIAN KANG[1,c]

[1]*Department of Biostatistics, University of Michigan,* [a]*gxma@umich.edu,* [b]*byzhao@umich.edu,* [c]*jiankang@umich.edu*
[2]*Department of Epidemiology,* [d]*hhabua@umich.edu*

In neuroimaging studies, it becomes increasingly important to study associations between different imaging modalities using image-on-image regression (IIR), which faces challenges in interpretation, statistical inference and prediction. Our motivating problem is how to predict task-evoked fMRI activity using resting-state fMRI data in the Human Connectome Project (HCP). The main difficulty lies in effectively combining different types of imaging predictors with varying resolutions and spatial domains in IIR. To address these issues, we develop Bayesian Image-on-image Regression via Deep Kernel Learning Gaussian Processes (BIRD-GP) and develop efficient posterior computation methods through Stein variational gradient descent. We demonstrate the advantages of BIRD-GP over state-of-the-art IIR methods using extensive simulations where we synthesize data based on MNIST, Fashion MNIST and fMRI data from HCP. For HCP data analysis using BIRD-GP, we combine the voxelwise fALFF maps and regionwise connectivity matrices to predict fMRI contrast maps for language and social recognition tasks. We show that fALFF is less predictive than the connectivity matrix for both tasks. Additionally, we identify features from the resting-state fMRI data that are important for task fMRI prediction.

REFERENCES

BÁEZ-MENDOZA, R. and SCHULTZ, W. (2013). The role of the striatum in social behavior. *Front. Neurosci.* **7** 233.

BINDER, J. R., GROSS, W. L., ALLENDORFER, J. B., BONILHA, L., CHAPIN, J., EDWARDS, J. C., GRABOWSKI, T. J., LANGFITT, J. T., LORING, D. W. et al. (2011). Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study. *NeuroImage* **54** 1465–1475.

BURGALETA, M., SANJUÁN, A., VENTURA-CAMPOS, N., SEBASTIAN-GALLES, N. and ÁVILA, C. (2016). Bilingualism at the core of the brain. Structural differences between bilinguals and monolinguals revealed by subcortical shape analysis. *NeuroImage* **125** 437–445.

CALLEJAS, A., SHULMAN, G. L. and CORBETTA, M. (2014). Dorsal and ventral attention systems underlie social and symbolic cueing. *J. Cogn. Neurosci.* **26** 63–80.

CAPOZZI, F. and RISTIC, J. (2018). How attention gates social interactions. *Ann. N.Y. Acad. Sci.* **1426** 179–198.

CARVALHO, D. V., PEREIRA, E. M. and CARDOSO, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics* **8** 1–34. https://doi.org/10.3390/electronics8080832

CASTELLI, F., HAPPÉ, F., FRITH, U. and FRITH, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage* **12** 314–325.

DUBOIS, J. and ADOLPHS, R. (2016). Building a science of individual differences from fMRI. *Trends Cogn. Sci.* **20** 425–443.

DWORKIN, J. D., SWEENEY, E. M., SCHINDLER, M. K., CHAHIN, S., REICH, D. S. and SHINOHARA, R. T. (2016). PREVAIL: Predicting recovery through estimation and visualization of active and incident lesions. *NeuroImage Clin.* **12** 293–299.

EGOROVA, N., VELDSMAN, M., CUMMING, T. and BRODTMANN, A. (2017). Fractional amplitude of low-frequency fluctuations (fALFF) in post-stroke depression. *NeuroImage Clin.* **16** 116–124.

EVANS, A. C., COLLINS, D. L., MILLS, S., BROWN, E. D., KELLY, R. L. and PETERS, T. M. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. In 1993 *IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference* 1813–1817. IEEE.

Key words and phrases. Bayesian deep learning, human connectome project, neuroimaging.

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. MR1995715 https://doi.org/10.1198/016214503000170

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.

GORDON, E. M., LAUMANN, T. O., GILMORE, A. W., NEWBOLD, D. J., GREENE, D. J., BERG, J. J., ORTEGA, M., HOYT-DRAZEN, C., GRATTON, C. et al. (2017). Precision functional mapping of individual human brains. *Neuron* **95** 791–807.

GUO, C., KANG, J. and JOHNSON, T. D. (2022). A spatial Bayesian latent factor model for image-on-image regression. *Biometrics* **78** 72–84. MR4408571 https://doi.org/10.1111/biom.13420

HÄRKÖNEN, E., HERTZMANN, A., LEHTINEN, J. and PARIS, S. (2020). GANSpace: Discovering interpretable GAN controls. In *Advances in Neural Information Processing Systems* **33** 9841–9850.

HARREWIJN, A., ABEND, R., LINKE, J., BROTMAN, M. A., FOX, N. A., LEIBENLUFT, E., WINKLER, A. M. and PINE, D. S. (2020). Combining fMRI during resting state and an attention bias task in children. *NeuroImage* **205** 116301. https://doi.org/10.1016/j.neuroimage.2019.116301

HUANG, H., YU, P. S. and WANG, C. (2018). An introduction to image synthesis with generative adversarial nets. *CoRR*. Preprint. Available at arXiv:1803.04469.

ISOLA, P., ZHU, J.-Y., ZHOU, T. and EFROS, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1125–1134.

KETTELER, D., KASTRAU, F., VOHN, R. and HUBER, W. (2008). The subcortical role of language processing. High level linguistic features such as ambiguity-resolution and the human brain; an fMRI study. *NeuroImage* **39** 2002–2009.

KINGMA, D. P. and BA, J. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations* (*ICLR*).

LECUN, Y., CORTES, C. and BURGES, C. (2010). MNIST handwritten digit database. *ATT Labs* [Online]. Available at http://yann.lecun.com/exdb/mnist.

LINDQUIST, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. MR2530545 https://doi.org/10.1214/09-STS282

LIU, Q. and WANG, D. (2016). Stein Variational Gradient Descent: A general purpose Bayesian inference algorithm. In *Advances in Neural Information Processing Systems* (D. Lee, M. Sugiyama, U. Luxburg, I. Guyon and R. Garnett, eds.) **29**. Curran Associates, Red Hook.

MA, G., ZHAO, B., ABU-AMARA, H. and KANG, J. (2026). Supplement to "Bayesian image-on-image regression via deep kernel learning based Gaussian processes." https://doi.org/10.1214/25-AOAS2108SUPPA, https://doi.org/10.1214/25-AOAS2108SUPPB

MCNAB, F. and KLINGBERG, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nat. Neurosci.* **11** 103–107.

MORRIS, J. S., BALADANDAYUTHAPANI, V., HERRICK, R. C., SANNA, P. and GUTSTEIN, H. (2011). Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data. *Ann. Appl. Stat.* **5** 894–923. MR2840180 https://doi.org/10.1214/10-AOAS407

NGO, G. H., KHOSLA, M., JAMISON, K., KUCEYESKI, A. and SABUNCU, M. R. (2022). Predicting individual task contrasts from resting-state functional connectivity using a surface-based convolutional network. *NeuroImage* **248** 118849.

OFAN, R. H. and ZOHARY, E. (2007). Visual cortex activation in bilingual blind individuals during use of native and second language. *Cereb. Cortex* **17** 1249–1259.

PITCHER, D. and UNGERLEIDER, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends Cogn. Sci.* **25** 100–110.

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.

POWER, J. D., SCHLAGGAR, B. L. and PETERSEN, S. E. (2014). Studying brain organization via spontaneous fMRI signal. *Neuron* **84** 681–696.

RANJAN, A. and SINGH, V. P. (2025). Language processing in the brain: An fMRI study. In *Advances in Computers* **136** 493–564. Elsevier, Amsterdam.

SANTHANAM, V., MORARIU, V. I. and DAVIS, L. S. (2017). Generalized deep image to image regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 5609–5619.

SEYDELL-GREENWALD, A., WANG, X., NEWPORT, E. L., BI, Y. and STRIEM-AMIT, E. (2023). Spoken language processing activates the primary visual cortex. *PLoS ONE* **18** e0289671.

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** 1900–1911.

SRIPADA, C., ANGSTADT, M., RUTHERFORD, S., TAXALI, A. and SHEDDEN, K. (2020). Toward a "treadmill test" for cognition: Improved prediction of general cognitive ability from the task activated brain. *Hum. Brain Mapp.* **41** 3186–3197.

SRIVASTAVA, N., HINTON, G., KRIZHEVSKY, A., SUTSKEVER, I. and SALAKHUTDINOV, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15** 1929–1958. MR3231592

TANIMIZU, T., KENNEY, J. W., OKANO, E., KADOMA, K., FRANKLAND, P. W. and KIDA, S. (2017). Functional connectivity of multiple brain regions required for the consolidation of social recognition memory. *J. Neurosci.* **37** 4103–4116.

TAVOR, I., JONES, O. P., MARS, R., SMITH, S., BEHRENS, T. and JBABDI, S. (2016). Task-free MRI predicts individual differences in brain activity during task performance. *Science* **352** 216–220.

VAN ESSEN, D. C., UGURBIL, K., AUERBACH, E., BARCH, D., BEHRENS, T. E., BUCHOLZ, R., CHANG, A., CHEN, L., CORBETTA, M. et al. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage* **62** 2222–2231.

VARRIER, R. S. and FINN, E. S. (2022). Seeing social: A neural signature for conscious perception of social interactions. *J. Neurosci.* **42** 9211–9226.

WANG, K., LEOPOLD, D. R., BANICH, M. T., REINEBERG, A. E., WILLCUTT, E. G., CUTTING, L. E., DEL TUFO, S. N., THOMPSON, L. A., OPFER, J. et al. (2019). Characterizing and decomposing the neural correlates of individual differences in reading ability among adolescents with task-based fMRI. *Dev. Cogn. Neurosci.* **37** 100647.

WANG, X., KRIEGER-REDWOOD, K., ZHANG, M., CUI, Z., WANG, X., KARAPANAGIOTIDIS, T., DU, Y., LEECH, R., BERNHARDT, B. C. et al. (2023). Physical distance to sensory-motor landmarks predicts language function. *Cereb. Cortex* **33** 4305–4318.

WHEATLEY, T., MILLEVILLE, S. C. and MARTIN, A. (2007). Understanding animate agents: Distinct roles for the social network and mirror system. *Psychol. Sci.* **18** 469–474.

WU-MINN HCP (2017). WU-Minn HCP 1200 subjects data release: Reference manual.

XIAO, H., RASUL, K. and VOLLGRAF, R. (2017). Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. Preprint. Available at arXiv:1708.07747.

ZHANG, Q., WU, Y. N. and ZHU, S. C. (2018). Interpretable convolutional neural networks. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 8827–8836.

ZHAO, Y., KANG, J. and LONG, Q. (2015). Bayesian multiresolution variable selection for ultra-high dimensional neuroimaging data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **15** 537–550.

ZOU, Q.-H., ZHU, C.-Z., YANG, Y., ZUO, X.-N., LONG, X.-Y., CAO, Q.-J., WANG, Y.-F. and ZANG, Y.-F. (2008). An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF. *J. Neurosci. Methods* **172** 137–141.

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 560–577
https://doi.org/10.1214/25-AOAS2119

IDENTIFICATION OF GENETIC FACTORS ASSOCIATED WITH CORPUS CALLOSUM MORPHOLOGY: CONDITIONAL STRONG INDEPENDENCE SCREENING FOR NON-EUCLIDEAN RESPONSES

BY ZHE GAO[1,a], JIN ZHU[2,c], YUE HU[3,d], WENLIANG PAN[4,e] AND XUEQIN WANG[1,b]

[1]*School of Management, University of Science and Technology of China*, [a]*gaozh8@mail.ustc.edu.cn*, [b]*wangxq20@ustc.edu.cn*
[2]*School of Mathematics, University of Birmingham*, [c]*j.zhu.7@bham.ac.uk*
[3]*School of Public Health, Yale University*, [d]*yue.hu@yale.edu*
[4]*State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences*, [e]*panwliang@amss.ac.cn*

The corpus callosum, the largest white matter structure in the brain, plays a critical role in interhemispheric communication. Variations in its morphology are associated with various neurological and psychological conditions, making it a key focus in neurogenetics. Age is known to influence the structure and morphology of the corpus callosum significantly, complicating the identification of specific genetic factors that contribute to its shape and size. We propose a conditional strong independence screening method to address these challenges for ultrahigh-dimensional predictors and non-Euclidean responses, incorporating prior knowledge such as age through a novel concept of conditional metric dependence, which quantifies nonlinear conditional dependencies among random objects in metric spaces without relying on predefined models. We apply this framework to identify genetic factors associated with the morphology of the corpus callosum. Simulation results demonstrate the efficacy of this method across various non-Euclidean data types, highlighting its potential to drive genetic discovery in neuroscience.

REFERENCES

BACHMAN, A. H., LEE, S. H., SIDTIS, J. J. and ARDEKANI, B. A. (2014). Corpus callosum shape and size changes in early Alzheimer's disease: A longitudinal MRI study using the OASIS brain database. *J. Alzheimer's Dis.* **39** 71–78.

BARUT, E., FAN, J. and VERHASSELT, A. (2016). Conditional sure independence screening. *J. Amer. Statist. Assoc.* **111** 1266–1277. MR3561948 https://doi.org/10.1080/01621459.2015.1092974

BISWAL, B. B., MENNES, M., ZUO, X.-N., GOHEL, S., KELLY, C., SMITH, S. M., BECKMANN, C. F., ADELSTEIN, J. S., BUCKNER, R. L. et al. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. USA* **107** 4734–4739. https://doi.org/10.1073/pnas.0911855107

CHEN, X., ZHANG, Y., CHEN, X. and LIU, Y. (2019). A simple model-free survival conditional feature screening. *Statist. Probab. Lett.* **146** 156–160. MR3882343 https://doi.org/10.1016/j.spl.2018.11.019

COLES, C. H., SHEN, Y., TENNEY, A. P., SIEBOLD, C., SUTTON, G. C., LU, W., GALLAGHER, J. T., JONES, E. Y., FLANAGAN, J. G. et al. (2011). Proteoglycan-specific molecular switch for RPTPσ clustering and neuronal extension. *Science* **332** 484–488.

CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 463–482. MR3611755 https://doi.org/10.1111/rssb.12169

DRYDEN, I. L. and MARDIA, K. V. (2016). *Statistical Shape Analysis with Applications in R*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3559734 https://doi.org/10.1002/9781119072492

ECCHER, M. (2014). Corpus callosum. In *Encyclopedia of the Neurological Sciences*, 2nd ed. (M. J. Aminoff and R. B. Daroff, eds.) 867–868. Academic Press, Oxford. https://doi.org/10.1016/B978-0-12-385157-4.01137-4

EDWARDS, T. J., SHERR, E. H., BARKOVICH, A. J. and RICHARDS, L. J. (2014). Clinical, genetic and imaging findings identify new causes for corpus callosum development syndromes. *Brain* **137** 1579–1613. https://doi.org/10.1093/brain/awt358

Key words and phrases. Conditional metric dependence, conditional strong independence screening, non-Euclidean response, ultrahigh-dimensional data analysis.

FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x

FEBRERO-BANDE, M. and DE LA FUENTE, M. (2012). Statistical computing in functional data analysis: The R package fda.usc. *J. Stat. Softw.* **51** 1–28. https://doi.org/10.18637/jss.v051.i04

FEDERER, H. (2014). *Geometric Measure Theory*. Springer, Berlin.

FLETCHER, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vis.* **105** 171–185. MR3104017 https://doi.org/10.1007/s11263-012-0591-y

FUKUMIZU, K., GRETTON, A., SUN, X. and SCHÖLKOPF, B. (2008). Kernel measures of conditional dependence. In *Advances in Neural Information Processing Systems* 489–496.

GAO, Z., ZHU, J., HU, Y., PAN, W. and WANG, X. (2026). Supplement to “Identification of Genetic Factors Associated with Corpus Callosum Morphology: Conditional Strong Independence Screening for Non-Euclidean Responses.” https://doi.org/10.1214/25-AOAS2119SUPPA, https://doi.org/10.1214/25-AOAS2119SUPPB

HINKLE, J., FLETCHER, P. T. and JOSHI, S. (2014). Intrinsic polynomials for regression on Riemannian manifolds. *J. Math. Imaging Vision* **50** 32–52. MR3233133 https://doi.org/10.1007/s10851-013-0489-5

HONG, H. G., KANG, J. and LI, Y. (2018). Conditional screening for ultra-high dimensional covariates with survival outcomes. *Lifetime Data Anal.* **24** 45–71. MR3742906 https://doi.org/10.1007/s10985-016-9387-7

HONG, H. G., WANG, L. and HE, X. (2016). A data-driven approach to conditional screening of high-dimensional variables. *Stat* **5** 200–212. MR3530327 https://doi.org/10.1002/sta4.115

HSU, M., DEDHIA, M., CRUSIO, W. and DELPRATO, A. (2019). Sex differences in gene expression patterns associated with the APOE4 allele. *F1000Res.* **8** 387. https://doi.org/10.12688/f1000research.18671.2

HU, Q. and LIN, L. (2017). Conditional sure independence screening by conditional marginal empirical likelihood. *Ann. Inst. Statist. Math.* **69** 63–96. MR3590712 https://doi.org/10.1007/s10463-015-0534-9

HU, W., HUANG, M., PAN, W., WANG, X., WEN, C., TIAN, Y., ZHANG, H. and ZHU, J. (2019). cdcsis: Conditional Distance Correlation Based Feature Screening and Conditional Independence Inference R package version 2.0.3.

HUANG, D., LI, R. and WANG, H. (2014). Feature screening for ultrahigh dimensional categorical data with applications. *J. Bus. Econom. Statist.* **32** 237–244. MR3207836 https://doi.org/10.1080/07350015.2013.863158

JOSHI, S. H., NARR, K. L., PHILIPS, O. R., NUECHTERLEIN, K. H., ASARNOW, R. F., TOGA, A. W. and WOODS, R. P. (2013). Statistical shape analysis of the corpus callosum in schizophrenia. *NeuroImage* **64** 547–559.

KENNEDY, H., VAN ESSEN, D. C. and CHRISTEN, Y. (2016). *Micro-, Meso-and Macro-Connectomics of the Brain*. Springer, Berlin.

KOIZUMI, K., HATTORI, Y., AHN, S. J., BUENDIA, I., CIACCIARELLI, A., UEKAWA, K., WANG, G., HILLER, A., ZHAO, L. et al. (2018). Apoε4 disrupts neurovascular regulation and undermines white matter integrity and cognitive function. *Nat. Commun.* **9** 1–11.

LI, R., ZHONG, W. and ZHU, L. (2012). Feature screening via distance correlation learning. *J. Amer. Statist. Assoc.* **107** 1129–1139. MR3010900 https://doi.org/10.1080/01621459.2012.695654

LIN, L. and SUN, J. (2016). Adaptive conditional feature screening. *Comput. Statist. Data Anal.* **94** 287–301. MR3412826 https://doi.org/10.1016/j.csda.2015.09.002

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 CHEN, X. (2018). Quantile screening for ultra-high-dimensional heterogeneous data conditional on some variables. *J. Stat. Comput. Simul.* **88** 329–342. MR3740732 https://doi.org/10.1080/00949655.2017.1389944

LU, S., CHEN, X. and WANG, H. (2021). Conditional distance correlation sure independence screening for ultra-high dimensional survival data. *Comm. Statist. Theory Methods* **50** 1936–1953. MR4241930 https://doi.org/10.1080/03610926.2019.1657454

LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883 https://doi.org/10.1214/12-AOP803

MAZARAKIS, N., MICHALOVICH, D., KARIS, A., GROSVELD, F. and GALJART, N. (1996). Zfp-37Is a member of the KRAB zinc finger gene family and is expressed in neurons of the developing and adult CNS. *Genomics* **33** 247–257. https://doi.org/10.1006/geno.1996.0189

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

PAN, W., WANG, X., XIAO, W. and ZHU, H. (2019). A generic sure independence screening procedure. *J. Amer. Statist. Assoc.* **114** 928–937. MR3963192 https://doi.org/10.1080/01621459.2018.1462709

PAN, W., WANG, X., ZHANG, H., ZHU, H. and ZHU, J. (2020). Ball covariance: A generic measure of dependence in Banach space. *J. Amer. Statist. Assoc.* **115** 307–317. MR4078465 https://doi.org/10.1080/01621459.2018.1543600

PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624

PULIDO, R., SERRA-PAGES, C., TANG, M. and STREULI, M. (1995). The LAR/PTP delta/PTP sigma subfamily of transmembrane protein-tyrosine-phosphatases: Multiple human LAR, PTP delta, and PTP sigma isoforms are expressed in a tissue-specific manner and associate with the LAR-interacting protein LIP. 1. *Proc. Natl. Acad. Sci. USA* **92** 11686–11690.

SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538. MR4124333 https://doi.org/10.1214/19-AOS1857

TANAKA-ARAKAWA, M. M., MATSUI, M., TANAKA, C., UEMATSU, A., UDA, S., MIURA, K., SAKAI, T. and NOGUCHI, K. (2015). Developmental changes in the corpus callosum from infancy to early adulthood: A structural magnetic resonance imaging study. *PLoS ONE* **10** e0118760.

VERMEULEN, C. L., DU TOIT, P. J., VENTER, G. and HUMAN-BARON, R. (2023). A morphological study of the shape of the corpus callosum in normal, schizophrenic and bipolar patients. *J. Anat.* **242** 153–163.

WANG, X., PAN, W., HU, W., TIAN, Y. and ZHANG, H. (2015). Conditional distance correlation. *J. Amer. Statist. Assoc.* **110** 1726–1734. MR3449068 https://doi.org/10.1080/01621459.2014.993081

WANG, X., ZHU, J., PAN, W., ZHU, J. and ZHANG, H. (2024). Nonparametric statistical inference via metric distribution function in metric spaces. *J. Amer. Statist. Assoc.* **119** 2772–2784. MR4833914 https://doi.org/10.1080/01621459.2023.2277417

WEN, C., PAN, W., HUANG, M. and WANG, X. (2018). Sure independence screening adjusted for confounding covariates with ultrahigh dimensional data. *Statist. Sinica* **28** 293–317. MR3752262

XUE, J. and LIANG, F. (2017). A robust model-free feature screening method for ultrahigh-dimensional data. *J. Comput. Graph. Statist.* **26** 803–813. MR3765345 https://doi.org/10.1080/10618600.2017.1328364

ZHANG, S., PAN, J. and ZHOU, Y. (2018). Robust conditional nonparametric independence screening for ultrahigh-dimensional data. *Statist. Probab. Lett.* **143** 95–101. MR3854211 https://doi.org/10.1016/j.spl.2018.08.003

ZHENG, Q., HONG, H. G. and LI, Y. (2020). Building generalized linear models with ultrahigh dimensional features: A sequentially conditional approach. *Biometrics* **76** 47–60. MR4098543 https://doi.org/10.1111/biom.13122

ZHU, J., PAN, W., ZHENG, W. and WANG, X. (2021). Ball: An R package for detecting distribution difference and association in metric spaces. *J. Stat. Softw.* **97** 1–31. https://doi.org/10.18637/jss.v097.i06

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 578–603
https://doi.org/10.1214/25-AOAS2124
© Institute of Mathematical Statistics, 2026

SCALABLE MAGNETIC RESONANCE FINGERPRINTING: INCREMENTAL INFERENCE OF HIGH-DIMENSIONAL ELLIPTICAL MIXTURES FROM LARGE DATA VOLUMES

BY GEOFFROY OUDOUMANESSAH[1,2,3,a], THOMAS COUDERT[2,c], CAROLE LARTIZIEN[3,f], MICHEL DOJAT[1,2,d], THOMAS CHRISTEN[2,e] AND FLORENCE FORBES[1,b]

[1] *Univiversité Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK,* [a]*geoffroy.oudoumanessah@inria.fr,* [b]*florence.forbes@inria.fr*
[2] *Univiversité Grenoble Alpes, Inserm U1216, CHU Grenoble Alpes, Grenoble Institut des Neurosciences,* [c]*thomas.coudert@inserm.fr,* [d]*michel.dojat@inserm.fr,* [e]*thomas.christen@univ-grenoble-alpes.fr*
[3] *Univiversité Lyon, CNRS, Inserm, INSA Lyon, UCBL, CREATIS, UMR5220, U1294, F-69621,* [f]*carole.lartizien@creatis.insa-lyon.fr*

Magnetic Resonance Fingerprinting (MRF) is an emerging technology with the potential to revolutionize radiology and medical diagnostics. In comparison to traditional magnetic resonance imaging (MRI), MRF enables the rapid, simultaneous, noninvasive acquisition and reconstruction of multiple tissue parameters, paving the way for novel diagnostic techniques. In the original *matching* approach, reconstruction is based on the search for the best matches between in vivo acquired signals and a dictionary of high-dimensional simulated signals (fingerprints) with known tissue properties. A critical and limiting challenge is that the size of the simulated dictionary increases exponentially with the number of parameters, leading to an extremely costly matching. In this work we propose to address this scalability issue by considering probabilistic mixtures of high-dimensional elliptical distributions to learn more efficient dictionary representations. Mixture components are modelled as flexible elliptical shapes in low-dimensional subspaces. They are exploited to cluster similar signals and reduce their dimension locally clusterwise limiting information loss. To estimate such a mixture model, we provide a new incremental algorithm capable of handling large numbers of signals, allowing us to go far beyond the hardware limitations encountered by standard implementations. We demonstrate, on simulated and real data, that our method effectively manages large volumes of MRF data with maintained accuracy. It offers a more efficient solution for accurate tissue characterization and significantly reduces the computational burden, making the clinical application of MRF more practical and accessible.

REFERENCES

ARCHAMBEAU, C. and BACH, F. R. (2008). Sparse probabilistic projections. In 21*st International Conference on Neural Information Processing Systems* 73–80.

ARCHAMBEAU, C., DELANNAY, N. and VERLEYSEN, M. (2006). Robust probabilistic projections. In *Proceedings of the* 23*rd International Conference on Machine Learning* 33–40.

ARCHAMBEAU, C., DELANNAY, N. and VERLEYSEN, M. (2008). Mixtures of robust probabilistic principal component analyzers. *Neurocomputing* **71** 1274–1282.

BAEK, J. and MCLACHLAN, G. J. (2011). Mixtures of common t-factor analyzers for clustering high-dimensional microarray data. *Bioinformatics* **27** 1269–1276.

BALZANO, L., CHI, Y. and LU, Y. M. (2018). Streaming PCA and subspace tracking: The missing data case. In *Proceedings of the IEEE* **106** 1293–1310.

BARRIER, A., COUDERT, T., DELPHIN, A., LEMASSON, B. and CHRISTEN, T. (2024). MARVEL: MR fingerprinting with additional micRoVascular estimates using bidirectional LSTMs. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 259–269. Springer, Berlin.

Key words and phrases. Dimension reduction, clustering, incremental learning, high-dimensional mixture models, elliptical distributions, expectation maximization algorithm.

BELLAS, A., BOUVEYRON, C., COTTRELL, M. and LACAILLE, J. (2013). Model-based clustering of high-dimensional data streams with online mixture of probabilistic PCA. *Adv. Data Anal. Classif.* **7** 281–300. MR3103967 https://doi.org/10.1007/s11634-013-0133-7

BIPIN MEHTA, B., COPPO, S., FRANCES MCGIVNEY, D., IAN HAMILTON, J., CHEN, Y., JIANG, Y., MA, D., SEIBERLICH, N., GULANI, V. et al. (2019). Magnetic resonance fingerprinting: A technical review. *Magn. Reson. Med.* **81** 25–46.

BJØRNERUD, A. and EMBLEM, K. E. (2010). A fully automated method for quantitative cerebral hemodynamic analysis using DSC–MRI. *J. Cereb. Blood Flow Metab.* **30** 1066–1078.

BORKAR, V. S. (2009). *Stochastic Approximation: A Dynamical Systems Viewpoint* **48**. Springer, Gurgaon.

BOUVEYRON, C. and BRUNET-SAUMARD, C. (2014). Model-based clustering of high-dimensional data: A review. *Comput. Statist. Data Anal.* **71** 52–78. MR3131954 https://doi.org/10.1016/j.csda.2012.12.008

BOUVEYRON, C., GIRARD, S. and SCHMID, C. (2007). High-dimensional data clustering. *Comput. Statist. Data Anal.* **52** 502–519. MR2409998 https://doi.org/10.1016/j.csda.2007.02.009

BOUX, F., FORBES, F., ARBEL, J., LEMASSON, B. and BARBIER, E. L. (2021). Bayesian inverse regression for vascular magnetic resonance fingerprinting. *IEEE Trans. Med. Imag.* **40** 1827–1837.

BRADBURY, J., FROSTIG, R., HAWKINS, P., JOHNSON, M., LEARY, C., MACLAURIN, D., NECULA, G., PASZKE, A., VANDERPLAS, J. et al. (2018). JAX: Composable transformations of Python+NumPy programs. http://github.com/google/jax.

CABINI, R. F., BARZAGHI, L., CICOLARI, D., AROSIO, P., CARRAZZA, S., FIGINI, S., FILIBIAN, M., GAZZANO, A., KRAUSE, R. et al. (2024). Fast deep learning reconstruction techniques for preclinical magnetic resonance fingerprinting. *NMR Biomed.* **37** e5028.

CAMBANIS, S., HUANG, S. and SIMONS, G. (1981). On the theory of elliptically contoured distributions. *J. Multivariate Anal.* **11** 368–385. MR0629795 https://doi.org/10.1016/0047-259X(81)90082-8

CAPPÉ, O. and MOULINES, E. (2009). On-line expectation-maximization algorithm for latent data models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 593–613. MR2749909 https://doi.org/10.1111/j.1467-9868.2009.00698.x

CAULEY, S. F., SETSOMPOP, K., MA, D., JIANG, Y., YE, H., ADALSTEINSSON, E., GRISWOLD, M. A. and WALD, L. L. (2015). Fast group matching for MR fingerprinting reconstruction. *Magn. Reson. Med.* **74** 523–528.

CHRISTEN, T., BOLAR, D. S. and ZAHARCHUK, G. (2013). Imaging brain oxygenation with MRI using blood oxygenation approaches: Methods, validation, and clinical applications. *Amer. J. Neuroradiol.* **34** 1113–1123.

CHRISTEN, T., PANNETIER, N., NI, W. W., QIU, D., MOSELEY, M. E., SCHUFF, N. and ZAHARCHUK, G. (2014). MR vascular fingerprinting: A new approach to compute cerebral blood volume, mean vessel radius, and oxygenation maps in the human brain. *NeuroImage* **89** 262–270.

COHEN, O., ZHU, B. and ROSEN, M. S. (2018). MR fingerprinting deep reconstruction network (DRONE). *Magn. Reson. Med.* **80** 885–894.

COIFMAN, R. R. and LAFON, S. (2006). Diffusion maps. *Appl. Comput. Harmon. Anal.* **21** 5–30. MR2238665 https://doi.org/10.1016/j.acha.2006.04.006

COUDERT, T., DELPHIN, A., BARRIER, A., BARBIER, E., LEMASSON, B., WARNKING, J. and CHRISTEN, T. (2025a). MR fingerprinting for imaging brain hemodynamics and oxygenation. *J. Magn. Reson. Imaging*.

COUDERT, T., DELPHIN, A., BARRIER, A., LEGRIS, L., WARNKING, J. M., LAMALLE, L., DONEVA, M., LEMASSON, B., BARBIER, E. L. et al. (2025b). Relaxometry and contrast-free cerebral microvascular quantification using balanced steady-state free precession MR fingerprinting. *Magn. Reson. Med.* **94** 302–316.

CRANMER, K., BREHMER, J. and LOUPPE, G. (2020). The frontier of simulation-based inference. *Proc. Natl. Acad. Sci. USA* **117** 30055–30062. MR4263287 https://doi.org/10.1073/pnas.1912789117

D'ASPREMONT, A., EL GHAOUI, L., JORDAN, M. I. and LANCKRIET, G. R. G. (2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Rev.* **49** 434–448. MR2353806 https://doi.org/10.1137/050645506

DELEFORGE, A., FORBES, F., BA, S. O. and HORAUD, R. (2015). Hyper-spectral image analysis with partially latent regression and spatial Markov dependencies. *IEEE J. Sel. Top. Signal Process.* **9** 1037–1048.

DELEFORGE, A., FORBES, F. and HORAUD, R. (2015). High-dimensional regression with Gaussian mixtures and partially-latent response variables. *Stat. Comput.* **25** 893–911. MR3375624 https://doi.org/10.1007/s11222-014-9461-5

DELPHIN, A., COUDERT, T., FAN, A., MOSELEY, M. E., ZAHARCHUK, G. and CHRISTEN, T. (2023). MR vascular fingerprinting with 3D realistic blood vessel structures and machine learning to assess oxygenation changes in human volunteers. In 2023 *ISMRM & ISMRT Annual Meeting & Exhibition*.

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, Methodol.* **39** 1–38. MR0501537

DIEULEVEUT, A., FORT, G., MOULINES, E. and WAI, H.-T. (2023). Stochastic approximation beyond gradient for signal processing and machine learning. *IEEE Trans. Signal Process.* **71** 3117–3148. MR4649660 https://doi.org/10.1109/tsp.2023.3301121

FEARNHEAD, P. and PRANGLE, D. (2012). Constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 419–474. MR2925370 https://doi.org/10.1111/j.1467-9868.2011.01010.x

FORBES, F., NGUYEN, H. D., NGUYEN, T. and ARBEL, J. (2022). Summary statistics and discrepancy measures for approximate Bayesian computation via surrogate posteriors. *Stat. Comput.* **32** 85. MR4491498 https://doi.org/10.1007/s11222-022-10155-6

FREITAS LOPES, H. and WEST, M. (2004). Bayesian model assessment in factor analysis. *Statist. Sinica* **14** 41–67. MR2036762

FORT, G., MOULINES, E. and WAI, H.-T. (2020). A stochastic path-integrated differential estimator expectation maximization algorithm. In *Proceedings of the* 34*th Conference on Neural Information Processing Systems* (*NeurIPS*).

GHAHRAMANI, Z. and BEAL, M. J. (1999). Variational inference for Bayesian mixtures of factor analysers. In *Proceedings of the* 13*th International Conference on Neural Information Processing Systems*. *NIPS'*99 449–455. MIT Press, Cambridge, MA.

GHAHRAMANI, Z. and HINTON, G. E. (1997). The EM algorithm for mixtures of factor analyzers Technical Report, Univ. Toronto.

GILMAN, K., HONG, D., FESSLER, J. A. and BALZANO, L. (2023). Streaming probabilistic PCA for missing data with heteroscedastic noise. arXiv preprint. Available at arXiv:2310.06277.

GIRAUD, C. (2014). *Introduction to High-Dimensional Statistics*. CRC Press, Boca Raton.

GOLBABAEE, M., CHEN, D., GÓMEZ, P. A., MENZEL, M. I. and DAVIES, M. E. (2019). Geometry of deep learning for magnetic resonance fingerprinting. In *ICASSP* 2019-2019 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) 7825–7829. IEEE.

GOMEZ, E., GOMEZ-VILLEGAS, M. and MARIN, J. (2008). Multivariate exponential power distributions as mixtures of normal distributions with Bayesian applications. *Comm. Statist. Theory Methods* **37**.

GÓMEZ-SÁNCHEZ-MANZANO, E., GÓMEZ-VILLEGAS, M. A. and MARÍN, J. M. (2006). Sequences of elliptical distributions and mixtures of normal distributions. *J. Multivariate Anal.* **97** 295–310. MR2234024 https://doi.org/10.1016/j.jmva.2005.03.008

GU, Y., PAN, Y., FANG, Z., MA, L., ZHU, Y., ANDROJNA, C., ZHONG, K., YU, X. and SHEN, D. (2024). Deep learning-assisted preclinical MR fingerprinting for sub-millimeter T1 and T2 mapping of entire macaque brain. *Magn. Reson. Med.* **91** 1149–1164.

GU, Y., WANG, C. Y., ANDERSON, C. E., LIU, Y., HU, H., JOHANSEN, M. L., MA, D., JIANG, Y., RAMOS-ESTEBANEZ, C. et al. (2018). Fast magnetic resonance fingerprinting for dynamic contrast-enhanced studies in mice. *Magn. Reson. Med.* **80** 2681–2690.

GUO, Y. and BONDELL, H. (2023). On robust probabilistic principal component analysis using multivariate t-distributions. *Comm. Statist. Theory Methods* **52** 8261–8279. MR4652672 https://doi.org/10.1080/03610926.2022.2060512

HÄGGSTRÖM, H., RODRIGUES, P. L., OUDOUMANESSAH, G., FORBES, F. and PICCHINI, U. (2024). Fast, accurate and lightweight sequential simulation-based inference using Gaussian locally linear mappings. arXiv preprint. Available at arXiv:2403.07454.

HALKO, N., MARTINSSON, P. G. and TROPP, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.* **53** 217–288. MR2806637 https://doi.org/10.1137/090771806

HARGREAVES, B. Bloch Equation Simulator. http://www-mrsrl.stanford.edu/~brian/blochsim/.

HONG, D., GILMAN, K., BALZANO, L. and FESSLER, J. A. (2021b). HePPCAT: Probabilistic PCA for data with heteroscedastic noise. *IEEE Trans. Signal Process.* **69** 4819–4834. MR4313206 https://doi.org/10.1109/TSP.2021.3104979

HONG, D., YANG, F., FESSLER, J. A. and BALZANO, L. (2023). Optimally weighted PCA for high-dimensional heteroscedastic data. *SIAM J. Math. Data Sci.* **5** 222–250. MR4567414 https://doi.org/10.1137/22M1470244

JOLLIFFE, I. T. and CADIMA, J. (2016). Principal component analysis: A review and recent developments. *Philos. Trans. Roy. Soc. A* **374** 20150202. MR3479904 https://doi.org/10.1098/rsta.2015.0202

KARIMI, B., MIASOJEDOW, B., MOULINES, E. and WAI, H.-T. (2019a). Non-asymptotic analysis of biased stochastic approximation scheme. *Proc. Mach. Learn. Res.* **99** 1–31.

KARIMI, B., WAI, H.-T., MOULINES, R. and LAVIELLE, M. (2019b). On the global convergence of (fast) incremental expectation maximization methods. In *Proceedings of the* 33*rd Conference on Neural Information Processing Systems* (*NeurIPS*).

KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā Ser. A* **32** 419–438. MR0287628

KÖRZDÖRFER, G., PFEUFFER, J., KLUGE, T., GEBHARDT, M., HENSEL, B., MEYER, C. H. and NITTKA, M. (2019). Effect of spiral undersampling patterns on FISP MRF parameter maps. *Magn. Reson. Imaging* **62** 174–180.

KOTZ, S., KOZUBOWSKI, T. and PODGORSKI, K. (2001). *The Laplace Distribution and Generalizations*: *A Revisit with Applications to Communications*, *Economics*, *Engineering*, *and Finance*. *Progress in Mathematics*. Birkhäuser, Boston, MA.

KOTZ, S. and NADARAJAH, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge Univ. Press, Cambridge. MR2038227 https://doi.org/10.1017/CBO9780511550683

KUGLER, B., FORBES, F. and DOUTÉ, S. (2022). Fast Bayesian inversion for high dimensional inverse problems. *Stat*. *Comput*. **32** 31. MR4402178 https://doi.org/10.1007/s11222-021-10019-5

KUHN, E., MATIAS, C. and REBAFKA, T. (2020). Properties of the stochastic approximation EM algorithm with mini-batch sampling. *Stat*. *Comput*. **30** 1725–1739. MR4156345 https://doi.org/10.1007/s11222-020-09968-0

LAGOGIANNIS, I., MEISSEN, F., KAISSIS, G. and RUECKERT, D. (2024). Unsupervised pathology detection: A deep dive into the state of the art. *IEEE Trans*. *Med*. *Imag*. **43** 241–252.

LANGE, K. and SINSHEIMER, J. S. (1993). Normal/independent distributions and their applications in robust regression. *J*. *Comput*. *Graph*. *Statist*. **2** 175–198. MR1272391 https://doi.org/10.2307/1390698

LANGE, K. L., LITTLE, R. J. A. and TAYLOR, J. M. G. (1989). Robust statistical modeling using the *t* distribution. *J*. *Amer*. *Statist*. *Assoc*. **84** 881–896. MR1134486

LAUAND, C. K. and MEYN, S. (2024). Revisiting Step-Size Assumptions in Stochastic Approximation.

LI, P. and HU, Y. (2023). Learned tensor low-CP-rank and Bloch response manifold priors for non-Cartesian MRF reconstruction. *IEEE Trans*. *Med*. *Imag*.

LI, P. and HU, Y. (2024). Deep magnetic resonance fingerprinting based on local and global vision transformer. *Med*. *Image Anal*. **95** 103198.

MA, D., GULANI, V., SEIBERLICH, N., LIU, K., SUNSHINE, J. L., DUERK, J. L. and GRISWOLD, M. A. (2013). Magnetic resonance fingerprinting. *Nature* **495** 187–192.

MAIRE, F., MOULINES, E. and LEFEBVRE, S. (2017). Online EM for functional data. *Comput*. *Statist*. *Data Anal*. **111** 27–47. MR3630216 https://doi.org/10.1016/j.csda.2017.01.006

MCGIVNEY, D. F., BOYACIOĞLU, R., JIANG, Y., POORMAN, M. E., SEIBERLICH, N., GULANI, V., KEENAN, K. E., GRISWOLD, M. A. and MA, D. (2020). Magnetic resonance fingerprinting review part 2: Technique and directions. *J*. *Magn*. *Reson*. *Imaging* **51** 993–1007.

MCGIVNEY, D. F., PIERRE, E., MA, D., JIANG, Y., SAYBASILI, H., GULANI, V. and GRISWOLD, M. A. (2014). SVD compression for magnetic resonance fingerprinting in the time domain. *IEEE Trans*. *Med*. *Imag*. **33** 2311–2322.

MCLACHLAN, G. J. and KRISHNAN, T. (2008). *The EM Algorithm and Extensions*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, Hoboken, NJ. MR2392878 https://doi.org/10.1002/9780470191613

MCLACHLAN, G. J., PEEL, D. and BEAN, R. W. (2003). Modelling high-dimensional data by mixtures of factor analyzers. *Comput*. *Statist*. *Data Anal*. **41** 379–388. MR1973720 https://doi.org/10.1016/S0167-9473(02)00183-4

MENG, X.-L. and RUBIN, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80** 267–278. MR1243503 https://doi.org/10.1093/biomet/80.2.267

MONGA, A., SINGH, D., DE MOURA, H. L., ZHANG, X., ZIBETTI, M. V. and REGATTE, R. R. (2024). Emerging trends in magnetic resonance fingerprinting for quantitative biomedical imaging applications: A review. *Bioengineering* **11** 236.

NERI, J., DEPALLE, P. and BADEAU, R. (2021). Approximate inference and learning of state space models with Laplace noise. *IEEE Trans*. *Signal Process*. **69** 3176–3189. MR4274015 https://doi.org/10.1109/TSP.2021.3075146

NGUYEN, H. D. and FORBES, F. (2022). Global implicit function theorems and the online expectation-maximisation algorithm. *Aust*. *N*. *Z*. *J*. *Stat*. **64** 255–281. MR4467060 https://doi.org/10.1111/anzs.12356

NGUYEN, H. D., FORBES, F. and MCLACHLAN, G. J. (2020). Mini-batch learning of exponential family finite mixture models. *Stat*. *Comput*. **30** 731–748. MR4108674 https://doi.org/10.1007/s11222-019-09919-4

OUDOUMANESSAH, G., COUDERT, T., MEYER, L., DELPHIN, A., CHRISTEN, T., DOJAT, M., LARTIZIEN, C. and FORBES, F. (2025). Cluster globally, reduce locally: Scalable efficient dictionary compression for magnetic resonance fingerprinting. In *International Symposium in Biomedical Imaging* 1–5.

OUDOUMANESSAH, G., COUDERT, T., LARTIZIEN, C., DOJAT, M., CHRISTEN, T. and FORBES, F. (2026). Supplement to "Scalable magnetic resonance fingerprinting: Incremental inference of high-dimensional elliptical mixtures from large data volumes." https://doi.org/10.1214/25-AOAS2124SUPPA, https://doi.org/10.1214/25-AOAS2124SUPPB

POORMAN, M. E., MARTIN, M. N., MA, D., MCGIVNEY, D. F., GULANI, V., GRISWOLD, M. A. and KEENAN, K. E. (2020). Magnetic resonance fingerprinting Part 1: Potential uses, current challenges, and recommendations. *J*. *Magn*. *Reson*. *Imaging* **51** 675–692.

RONNEBERGER, O., FISCHER, P. and BROX, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI* 2015: 18*th International Conference*, *Munich*, *Germany*, *October* 5-9, 2015, *Proceedings*, *Part III* **18** 234–241. Springer, Berlin.

SATOPAA, V., ALBRECHT, J., IRWIN, D. and RAGHAVAN, B. (2011). Finding a" kneedle" in a haystack: Detecting knee points in system behavior. In 2011 31*st International Conference on Distributed Computing Systems Workshops* 166–171. IEEE.

SCHOTT, J. R. (2017). *Matrix Analysis for Statistics*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3497549

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

SOYAK, R., NAVRUZ, E., ERSOY, E. O., CRUZ, G., PRIETO, C., KING, A. P., UNAY, D. and OKSUZ, I. (2021). Channel attention networks for robust MR fingerprint matching. *IEEE Trans*. *Biomed*. *Eng*. **69** 1398–1405.

TIPPAREDDY, C., ZHAO, W., SUNSHINE, J. L., GRISWOLD, M., MA, D. and BADVE, C. (2021). Magnetic resonance fingerprinting: An overview. *Eur*. *J*. *Nucl*. *Med*. *Mol*. *Imaging* **48** 4189–4200.

TIPPING, M. E. and BISHOP, C. M. (1999a). Probabilistic principal component analysis. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **61** 611–622. MR1707864 https://doi.org/10.1111/1467-9868.00196

TIPPING, M. E. and BISHOP, C. M. (1999b). Mixtures of probabilistic principal component analyzers. *Neural Comput*. **11** 443–482.

TIPPING, M. E. and BISHOP, C. M. (1999c). Probabilistic principal component analysis. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **61** 611–622. MR1707864 https://doi.org/10.1111/1467-9868.00196

TIPPING, M. E. and BISHOP, C. M. (1999d). Mixtures of probabilistic principal component analyzers. *Neural Comput*. **11** 443–482.

ULLAH, I., HASSAN, A. M., SAAD, R. M. and OMER, H. (2023). GPU accelerated grouped magnetic resonance fingerprinting using clustering techniques. *Magn*. *Reson*. *Imaging* **97** 13–23.

VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, Ł. and POLOSUKHIN, I. (2017). Attention is all you need. *Adv*. *Neural Inf*. *Process*. *Syst*. **30**.

WANG, C. Y., COPPO, S., MEHTA, B. B., SEIBERLICH, N., YU, X. and GRISWOLD, M. A. (2019a). Magnetic resonance fingerprinting with quadratic RF phase for measurement of T2* simultaneously with δf, T1, and T2. *Magn*. *Reson*. *Med*. **81** 1849–1862.

WANG, Z., BOVIK, A. C., SHEIKH, H. R. and SIMONCELLI, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Trans*. *Image Process*. **13** 600–612.

WANG, Z., ZHANG, J., CUI, D., XIE, J., LYU, M., HUI, E. S. and WU, E. X. (2019b). Magnetic resonance fingerprinting using a fast dictionary searching algorithm: MRF-ZOOM. *IEEE Trans*. *Biomed*. *Eng*. **66** 1526–1535.

WANSAPURA, J. P., HOLLAND, S. K., DUNN, R. S. and BALL, W. S. JR (1999). NMR relaxation times in the human brain at 3.0 tesla. *J*. *Magn*. *Reson*. *Imaging* **9** 531–538.

WEN, Z. and YIN, W. (2013). A feasible method for optimization with orthogonality constraints. *Math*. *Program*. **142** 397–434. MR3127080 https://doi.org/10.1007/s10107-012-0584-1

XU, A. S., BALZANO, L. and FESSLER, J. A. (2023). HeMPPCAT: Mixtures of probabilistic principal component analysers for data with heteroscedastic noise. In *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*) 1–5.

YANG, M., MA, D., JIANG, Y., HAMILTON, J., SEIBERLICH, N., GRISWOLD, M. A. and MCGIVNEY, D. (2018). Low rank approximation methods for MR fingerprinting with large scale dictionaries. *Magn*. *Reson*. *Med*. **79** 2392–2400.

YE, H., CAULEY, S. F., GAGOSKI, B., BILGIC, B., MA, D., JIANG, Y., DU, Y. P., GRISWOLD, M. A., WALD, L. L. et al. (2017). Simultaneous multislice magnetic resonance fingerprinting (SMS-MRF) with direct-spiral slice-GRAPPA (ds-SG) reconstruction. *Magn*. *Reson*. *Med*. **77** 1966–1974.

ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2006). Sparse principal component analysis. *J*. *Comput*. *Graph*. *Statist*. **15** 265–286. MR2252527 https://doi.org/10.1198/106186006X113430

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 604–622
https://doi.org/10.1214/25-AOAS2133

A GENERAL FRAMEWORK FOR INVESTIGATING NEURODEVELOPMENT OF BRAIN FUNCTIONAL NETWORKS USING MULTISITE AND LONGITUDINAL NEUROIMAGING

BY JOSHUA LUKEMIRE[a], YAOTIAN WANG[b] AND YING GUO[c]

Department of Biostatistics and Bioinformatics, Emory University, [a]*joshua.lukemire@emory.edu,* [b]*yaotian.wang@emory.edu,* [c]*yguo2@emory.edu*

In recent years, longitudinal, multisite imaging studies have emerged as key tools for investigating brain function. These studies follow a large number of participants for an extended period, offering exciting opportunities to uncover brain functional network changes over time as a function of clinical and demographic covariates. However, these studies also introduce many statistical challenges such as site-effects and accounting for the heterogeneous nature of network differences between subjects. Robust statistical methods are highly needed to address these issues, but to date, there has been little methods development addressing these problems in the context of data-driven brain network estimation. This work addresses this gap in the literature, introducing a general Bayesian framework, REMBRAiNDT, incorporating site- and subject-effects into the network decomposition, while also enabling covariate effect estimation and efficient information pooling across brain locations. We use our procedure to conduct a novel analysis of neurodevelopment among adolescents in the longitudinal, multisite ABCD study. We find extensive evidence of increasing functional integration with age in networks associated with higher order cognitive processes. Our study is one of the first to examine neurodevelopment using blind source separation in the longitudinal ABCD study data, and the findings enrich earlier cross-sectional results on neurodevelopment.

REFERENCES

ABROL, A., FU, Z., DU, Y., WILSON, T. W., WANG, Y.-P., STEPHEN, J. M. and CALHOUN, V. D. (2023). Developmental and aging resting functional magnetic resonance imaging brain state adaptations in adolescents and adults: A large N (> 47K) study. *Hum. Brain Mapp.* **44** 2158–2175.

ALLEN, E. A., ERHARDT, E. B., WEI, Y., EICHELE, T. and CALHOUN, V. D. (2012). Capturing inter-subject variability with group independent component analysis of fMRI data: A simulation study. *NeuroImage* **59** 4141–4159.

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

BEER, J. C., TUSTISON, N. J., COOK, P. A., DAVATZIKOS, C., SHELINE, Y. I., SHINOHARA, R. T., LINN, K. A., INITIATIVE, A. D. N. et al. (2020). Longitudinal combat: A method for harmonizing longitudinal multi-scanner imaging data. *NeuroImage* **220** 117129.

BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245 https://doi.org/10.1214/aos/1013699998

BIJSTERBOSCH, J. D., WOOLRICH, M. W., GLASSER, M. F., ROBINSON, E. C., BECKMANN, C. F., VAN ESSEN, D. C., HARRISON, S. J. and SMITH, S. M. (2018). The relationship between spatial configuration and functional connectivity of brain regions. *eLife* **7** e32992.

BISWAL, B., ZERRIN YETKIN, F., HAUGHTON, V. M. and HYDE, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* **34** 537–541.

BUCKNER, R. L., KRIENEN, F. M., CASTELLANOS, A., DIAZ, J. C. and YEO, B. T. (2011). The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106** 2322–2345.

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

Key words and phrases. Blind source separation, batch effects, neurodevelopment, functional connectivity, functional integration.

Carvalho, C. M., Polson, N. G. and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480. MR2650751 https://doi.org/10.1093/biomet/asq017

Casey, B. J., 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.

Chan, M. Y., Park, D. C., Savalia, N. K., Petersen, S. E. and Wig, G. S. (2014). Decreased segregation of brain systems across the healthy adult lifespan. *Proc. Natl. Acad. Sci. USA* **111** E4997–E5006.

Dansereau, C., Benhajali, Y., Risterucci, C., Pich, E. M., Orban, P., Arnold, D. and Bellec, P. (2017). Statistical power and prediction accuracy in multisite resting-state fMRI connectivity. *NeuroImage* **149** 220–232.

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

Farahibozorg, S.-R., Bijsterbosch, J. D., Gong, W., Jbabdi, S., Smith, S. M., Harrison, S. J. and Woolrich, M. W. (2021). Hierarchical modelling of functional brain networks in population and individuals from big fMRI data. *NeuroImage* **243** 118513.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics* 287–302. Academic Press, New York. MR0736538

Fiecas, M., Cribben, I., Bahktiari, R. and Cummine, J. (2017). A variance components model for statistical inference on functional connectivity networks. *NeuroImage* **149** 256–266.

Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., Roalf, D. R., Satterthwaite, T. D., Gur, R. C. et al. (2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage* **161** 149–170.

Franke, K. and Gaser, C. (2012). Longitudinal changes in individual BrainAGE in healthy aging, mild cognitive impairment, and Alzheimer's disease. *J. Gerontopsychol. Geriatr. Psychiatry* **25** 235–245.

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

Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., Nelson, S. M., Coalson, R. S., Snyder, A. Z. et al. (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron* **98** 439–452.

Greicius, M. D., Flores, B. H., Menon, V., Glover, G. H., Solvason, H. B., Kenna, H. et al. (2007). Resting-state functional connectivity in major depression: Abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biol. Psychiatry* **62** 429–437.

Guo, Y. (2011). A general probabilistic model for group independent component analysis and its estimation methods. *Biometrics* **67** 1532–1542. MR2872404 https://doi.org/10.1111/j.1541-0420.2011.01601.x

Guo, Y. and Tang, L. (2013). A hierarchical model for probabilistic independent component analysis of multi-subject fMRI studies. *Biometrics* **69** 970–981. MR3146792 https://doi.org/10.1111/biom.12068

Harms, M. P., Somerville, L. H., Ances, B. M., Andersson, J., Barch, D. M., Bastiani, M., Bookheimer, S. Y., Brown, T. B., Buckner, R. L. et al. (2018). Extending the human connectome project across ages: Imaging protocols for the lifespan development and aging projects. *NeuroImage* **183** 972–984.

Harrison, S. J., Woolrich, M. W., Robinson, E. C., Glasser, M. F., Beckmann, C. F., Jenkinson, M. and Smith, S. M. (2015). Large-scale probabilistic functional modes from resting state fMRI. *NeuroImage* **109** 217–231.

Hoff, P. D. (2009). Simulation of the matrix Bingham–von Mises–Fisher distribution, with applications to multivariate and relational data. *J. Comput. Graph. Statist.* **18** 438–456. MR2749840 https://doi.org/10.1198/jcgs.2009.07177

Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis* 46. Wiley, New York.

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

Jernigan, T. L., Brown, S. A. and Dowling, G. J. (2018). The adolescent brain cognitive development study. *J. Res. Adolesc.* **28** 154–156.

Johnson, W. E., Li, C. and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8** 118–127.

Kass, R. E., Carlin, B. P., Gelman, A. and Neal, R. M. (1998). Markov chain Monte Carlo in practice: A roundtable discussion. *Amer. Statist.* **52** 93–100. MR1628427 https://doi.org/10.2307/2685466

Liu, Z., Whitaker, K. J., Smith, S. M. and Nichols, T. E. (2022). Improved interpretability of brain-behavior CCA with domain-driven dimension reduction. *Front. Neurosci.* **16** 851827.

López-Vicente, M., Agcaoglu, O., Pérez-Crespo, L., Estévez-López, F., Heredia-Genestar, J. M., Mulder, R. H., Flournoy, J. C., van Duijvenvoorde, A. C., Güroğlu, B. et al. (2021). Developmental changes in dynamic functional connectivity from childhood into adolescence. *Front. Syst. Neurosci.* **15** 724805.

LUKEMIRE, J., PAGNONI, G. and GUO, Y. (2023). Sparse Bayesian modeling of hierarchical independent component analysis: Reliable estimation of individual differences in brain networks. *Biometrics* **79** 3599–3611. MR4680745 https://doi.org/10.1111/biom.13867

LUKEMIRE, J., WANG, Y. and GUO, Y. (2026). Supplement to "A General Framework for Investigating Neurodevelopment of Brain Functional Networks using Multisite and Longitudinal Neuroimaging." https://doi.org/10.1214/25-AOAS2133SUPP

LUKEMIRE, J., WANG, Y., VERMA, A. and GUO, Y. (2020). HINT: A hierarchical independent component analysis toolbox for investigating brain functional networks using neuroimaging data. *J. Neurosci. Methods* **108726**.

MA, S., CALHOUN, V. D., PHLYPO, R. and ADALI, T. (2014). Dynamic changes of spatial functional network connectivity in healthy individuals and schizophrenia patients using independent vector analysis. *NeuroImage* **90** 196–206.

MAKALIC, E. and SCHMIDT, D. F. (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Process. Lett.* **23** 179–182.

MEJIA, A. F., NEBEL, M. B., WANG, Y., CAFFO, B. S. and GUO, Y. (2020). Template independent component analysis: Targeted and reliable estimation of subject-level brain networks using big data population priors. *J. Amer. Statist. Assoc.* **115** 1151–1177. MR4143456 https://doi.org/10.1080/01621459.2019.1679638

MINKA, T. P. (2001). Automatic choice of dimensionality for PCA. *Adv. Neural Inf. Process. Syst.* **13** 598–604.

MÜLLER, P., ERKANLI, A. and WEST, M. (1996). Bayesian curve fitting using multivariate normal mixtures. *Biometrika* **83** 67–79. MR1399156 https://doi.org/10.1093/biomet/83.1.67

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

SATTERTHWAITE, T. D., WOLF, D. H., ERUS, G., RUPAREL, K., ELLIOTT, M. A., GENNATAS, E. D., HOPSON, R., JACKSON, C., PRABHAKARAN, K. et al. (2013). Functional maturation of the executive system during adolescence. *J. Neurosci.* **33** 16249–16261.

SCHAEFER, A., KONG, R., GORDON, E. M., LAUMANN, T. O., ZUO, X.-N., HOLMES, A. J., EICKHOFF, S. B. and YEO, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28** 3095–3114.

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

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.

SHI, R. and GUO, Y. (2016). Investigating differences in brain functional networks using hierarchical covariate-adjusted independent component analysis. *Ann. Appl. Stat.* **10** 1930–1957. MR3592043 https://doi.org/10.1214/16-AOAS946

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

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

THIJSSEN, S., COLLINS, P. F., WEISS, H. and LUCIANA, M. (2021). The longitudinal association between externalizing behavior and frontoamygdalar resting-state functional connectivity in late adolescence and young adulthood. *J. Child Psychol. Psychiatry* **62** 857–867.

TIAN, Y., MARGULIES, D. S., BREAKSPEAR, M. and ZALESKY, A. (2020). Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nat. Neurosci.* **23** 1421–1432.

VAN ESSEN, D. C., SMITH, S. M., BARCH, D. M., BEHRENS, T. E., YACOUB, E., UGURBIL, K., CONSORTIUM, W.-M. H. et al. (2013). The WU-Minn human connectome project: An overview. *NeuroImage* **80** 62–79.

VAN DER PAS, S., SZABÓ, B. and VAN DER VAART, A. (2017). Uncertainty quantification for the horseshoe (with discussion). *Bayesian Anal.* **12** 1221–1274. With a rejoinder by the authors. MR3724985 https://doi.org/10.1214/17-BA1065

WALKER, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Comm. Statist. Simulation Comput.* **36** 45–54. MR2370888 https://doi.org/10.1080/03610910601096262

WANG, Y. and GUO, Y. (2019). A hierarchical independent component analysis model for longitudinal neuroimaging studies. *NeuroImage* **189** 380–400.

WENDELKEN, C., FERRER, E., GHETTI, S., BAILEY, S. K., CUTTING, L. and BUNGE, S. A. (2017). Frontoparietal structural connectivity in childhood predicts development of functional connectivity and reasoning ability: A large-scale longitudinal investigation. *J. Neurosci.* **37** 8549–8558.

WENDELKEN, C., FERRER, E., WHITAKER, K. J. and BUNGE, S. A. (2016). Fronto-parietal network reconfiguration supports the development of reasoning ability. *Cereb. Cortex* **26** 2178–2190.

XIE, J., DOUGLAS, P. K., WU, Y. N., BRODY, A. L. and ANDERSON, A. E. (2017). Decoding the encoding of functional brain networks: An fMRI classification comparison of non-negative matrix factorization (NMF), independent component analysis (ICA), and sparse coding algorithms. *J. Neurosci. Methods* **282** 81–94.

YU, M., LINN, K. A., COOK, P. A., PHILLIPS, M. L., MCINNIS, M., FAVA, M., TRIVEDI, M. H., WEISSMAN, M. M., SHINOHARA, R. T. et al. (2018). Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Hum. Brain Mapp.* **39** 4213–4227.

ZHANG, W., LV, J., LI, X., ZHU, D., JIANG, X., ZHANG, S., ZHAO, Y., GUO, L., YE, J. et al. (2018). Experimental comparisons of sparse dictionary learning and independent component analysis for brain network inference from fMRI data. *IEEE Trans. Biomed. Eng.* **66** 289–299.

ZHU, T., HU, R., QIU, X., TAYLOR, M., TSO, Y., YIANNOUTSOS, C., NAVIA, B., MORI, S., EKHOLM, S. et al. (2011). Quantification of accuracy and precision of multi-center DTI measurements: A diffusion phantom and human brain study. *NeuroImage* **56** 1398–1411.

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 623–640
https://doi.org/10.1214/25-AOAS2134
© Institute of Mathematical Statistics, 2026

SEMIPARAMETRIC ANALYSIS OF INTERVAL-CENSORED DATA SUBJECT TO INACCURATE DIAGNOSES WITH A TERMINAL EVENT

BY YUHAO DENG[1,a], DONGLIN ZENG[1,b] AND YUANJIA WANG[2,c]

[1]*Department of Biostatistics, University of Michigan,* [a]*yuhaoden@umich.edu,* [b]*dzeng@umich.edu*
[2]*Department of Biostatistics, Columbia University,* [c]*yw2016@cumc.columbia.edu*

Interval-censoring frequently occurs in studies of chronic diseases where disease status is inferred from intermittently collected biomarkers. Although many methods have been developed to analyze such data, they typically assume perfect disease diagnosis, which often does not hold in practice due to the inherent imperfect clinical diagnosis of cognitive functions or measurement errors of biomarkers such as cerebrospinal fluid. In this work we introduce a semiparametric modeling framework using the Cox proportional hazards model to address interval-censored data in the presence of inaccurate disease diagnosis. Our model incorporates sensitivity and specificity of the diagnosis to account for uncertainty in whether the interval truly contains the disease onset. Furthermore, the framework accommodates scenarios involving a terminal event and when diagnosis is accurate, such as through postmortem analysis. We propose a nonparametric maximum likelihood estimation method for inference and develop an efficient EM algorithm to ensure computational feasibility. The regression coefficient estimators are shown to be asymptotically normal, achieving semiparametric efficiency bounds. We further validate our approach through extensive simulation studies and an application assessing Alzheimer's disease (AD) risk. We find that amyloid-beta is significantly associated with AD, but Tau is predictive of both AD and mortality.

REFERENCES

AREVALO-RODRIGUEZ, I., SMAILAGIC, N., I FIGULS, M. R., CIAPPONI, A., SANCHEZ-PEREZ, E., GIANNAKOU, A., PEDRAZA, O. L., COSP, X. B. and CULLUM, S. (2015). Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with Mild Cognitive Impairment (MCI). *Cochrane Database Syst. Rev.* **3**.

BEACH, T. G., MONSELL, S. E., PHILLIPS, L. E. and KUKULL, W. (2012). Accuracy of the clinical diagnosis of Alzheimer's disease at National Institute on Aging Alzheimer's Disease Centers, 2005–2010. *J. Neuropathol. Exp. Neurol.* **71** 266–273.

CHAPMAN, R. M., MAPSTONE, M., PORSTEINSSON, A. P., GARDNER, M. N., MCCRARY, J. W., DEGRUSH, E., REILLY, L. A., SANDOVAL, T. C. and GUILLILY, M. D. (2010). Diagnosis of Alzheimer's disease using neuropsychological testing improved by multivariate analyses. *J. Clin. Exp. Neuropsychol.* **32** 793–808.

CHOI, S. H., KIM, T. H., LIM, S., PARK, K. S., JANG, H. C. and CHO, N. H. (2011). Hemoglobin A1c as a diagnostic tool for diabetes screening and new-onset diabetes prediction: A 6-year community-based prospective study. *Diabetes Care* **34** 944–949.

CRANE, P. K., CARLE, A., GIBBONS, L. E., INSEL, P., MACKIN, R. S., GROSS, A., JONES, R. N., MUKHERJEE, S., CURTIS, S. M. et al. (2012). Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav.* **6** 502–516.

DENG, Y., ZENG, D. and WANG, Y. (2026). Supplement to "Semiparametric Analysis of Interval-Censored Data Subject to Inaccurate Diagnoses with a Terminal Event." https://doi.org/10.1214/25-AOAS2134SUPPA, https://doi.org/10.1214/25-AOAS2134SUPPB

FINE, J. P., JIANG, H. and CHAPPELL, R. (2001). On semi-competing risks data. *Biometrika* **88** 907–919. MR1872209 https://doi.org/10.1093/biomet/88.4.907

Key words and phrases. Cox proportional hazards model, interval censoring, diagnosis accuracy, terminal event, Alzheimer's disease.

GAO, F., ZENG, D., COUPER, D. and LIN, D. Y. (2019). Semiparametric regression analysis of multiple right- and interval-censored events. *J. Amer. Statist. Assoc.* **114** 1232–1240. MR4011775 https://doi.org/10.1080/01621459.2018.1482756

GAUGLER, J. E., KANE, R. L., JOHNSTON, J. A. and SARSOUR, K. (2013). Sensitivity and specificity of diagnostic accuracy in Alzheimer's disease: A synthesis of existing evidence. *Amer. J. Alzheimer's Dis. Other Dement.* **28** 337–347.

GOETGHEBEUR, E. and RYAN, L. (2000). Semiparametric regression analysis of interval-censored data. *Biometrics* **56** 1139–1144. MR1815593 https://doi.org/10.1111/j.0006-341X.2000.01139.x

HA, I. D., XIANG, L., PENG, M., JEONG, J.-H. and LEE, Y. (2020). Frailty modelling approaches for semi-competing risks data. *Lifetime Data Anal.* **26** 109–133. MR4048041 https://doi.org/10.1007/s10985-019-09464-2

HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24** 540–568. MR1394975 https://doi.org/10.1214/aos/1032894452

JIANG, F. and HANEUSE, S. (2017). A semi-parametric transformation frailty model for semi-competing risks survival data. *Scand. J. Stat.* **44** 112–129. MR3619697 https://doi.org/10.1111/sjos.12244

KIM, J. S. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 489–502. MR1983760 https://doi.org/10.1111/1467-9868.00398

KOOPERBERG, C. and CLARKSON, D. (1997). Hazard regression with interval-censored data. *Biometrics* **53** 1485–1494.

LEFFONDRÉ, K., TOURAINE, C., HELMER, C. and JOLY, P. (2013). Interval-censored time-to-event and competing risk with death: Is the illness-death model more accurate than the Cox model? *Int. J. Epidemiol.* **42** 1177–1186.

LINDSEY, J. C. and RYAN, L. M. (1998). Methods for interval-censored data. *Stat. Med.* **17** 219–238.

MITCHELL, A. J. (2009). A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *J. Psychiatr. Res.* **43** 411–431.

MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–465. With comments and a rejoinder by the authors. MR1803168 https://doi.org/10.2307/2669386

PETERSEN, R. C., AISEN, P. S., BECKETT, L. A., DONOHUE, M. C., GAMST, A. C., HARVEY, D. J., JACK, C. JR, JAGUST, W. J., SHAW, L. M. et al. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI) clinical characterization. *Neurology* **74** 201–209.

PIRES, M. C., COLOSIMO, E. A., VELOSO, G. A. and FERREIRA, R. D. S. B. (2021). Interval-censored data with misclassification: A Bayesian approach. *J. Appl. Stat.* **48** 907–923. MR4228779 https://doi.org/10.1080/02664763.2020.1753025

SELVIN, E., STEFFES, M. W., GREGG, E., BRANCATI, F. L. and CORESH, J. (2011). Performance of A1c for the classification and prediction of diabetes. *Diabetes Care* **34** 84–89.

SUN, J. (1997). Regression analysis of interval-censored failure time data. *Stat. Med.* **16** 497–504.

WANG, L., MCMAHAN, C. S., HUDGENS, M. G. and QURESHI, Z. P. (2016). A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics* **72** 222–231. MR3500591 https://doi.org/10.1111/biom.12389

WEI, Y., WOJTYŚ, M., SORRELL, L. and ROWE, P. (2023). Bivariate copula regression models for semi-competing risks. *Stat. Methods Med. Res.* **32** 1902–1918. MR4651764 https://doi.org/10.1177/09622802231188516

YANG, Z., RIZOPOULOS, D., HEIJNSDIJK, E. A., NEWCOMB, L. F. and ERLER, N. S. (2024). A Bayesian joint modelling for misclassified interval-censoring and competing risks. arXiv preprint. Available at arXiv:2404.09362.

ZENG, D., MAO, L. and LIN, D. Y. (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika* **103** 253–271. MR3509885 https://doi.org/10.1093/biomet/asw013

ZHANG, Y., HUA, L. and HUANG, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scand. J. Stat.* **37** 338–354. MR2682304 https://doi.org/10.1111/j.1467-9469.2009.00680.x

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 641–662
https://doi.org/10.1214/26-AOAS2150
© Institute of Mathematical Statistics, 2026

DYNAMIC CLASSIFICATION OF LATENT DISEASE PROGRESSION WITH AUXILIARY SURROGATE LABELS

BY ZEXI CAI[1,a], DONGLIN ZENG[2,c], KAREN S. MARDER[3,d], LAWRENCE S. HONIG[3,e] AND YUANJIA WANG[1,b]

[1]*Department of Biostatistics, Columbia University,* [a]*zc2626@columbia.edu,* [b]*yw2016@cumc.columbia.edu*
[2]*Department of Biostatistics, University of Michigan,* [c]*dzeng@umich.edu*
[3]*Department of Psychiatry, Columbia University Medical Center,* [d]*ksm1@cumc.columbia.edu,* [e]*lh456@cumc.columbia.edu*

Disease progression prediction based on patients' evolving health information is challenging when true disease states are unknown due to diagnostic capabilities or high costs. For example, the absence of gold-standard neurological diagnoses hinders distinguishing Alzheimer's disease (AD) from related conditions such as AD-related dementias (ADRDs), including Lewy body dementia (LBD). Combining temporally dependent surrogate labels and health markers may improve disease prediction. However, existing literature models informative surrogate labels and observed variables that reflect the underlying states using purely generative approaches, often posing unrealistic assumptions on the outcomes and suffering from misspecification thereof. We propose integrating the conventional hidden Markov model as a generative model with a time-varying discriminative classification model to simultaneously handle potentially misspecified surrogate labels and incorporate important markers of disease progression. We develop an adaptive forward-backward algorithm with subjective labels for estimation and utilize the modified posterior and Viterbi algorithms to predict the progression of future states or new patients based on objective markers only. Importantly, the adaptation eliminates the need to model the marginal distribution of longitudinal markers, a requirement in traditional algorithms. Asymptotic properties are established, and significant improvements in finite samples are demonstrated via simulation studies. Analysis of the neuropathological dataset of the National Alzheimer's Coordinating Center (NACC) shows much improved accuracy in distinguishing LBD from AD.

REFERENCES

BAKK, Z. and KUHA, J. (2018). Two-step estimation of models between latent classes and external variables. *Psychometrika* **83** 871–892. MR3875886 https://doi.org/10.1007/s11336-017-9592-7

BAKK, Z. and KUHA, J. (2021). Relating latent class membership to external variables: An overview. *Br. J. Math. Stat. Psychol.* **74** 340–362.

BENOIT, J. S., CHAN, W., LUO, S., YEH, H.-W. and DOODY, R. (2016). A hidden Markov model approach to analyze longitudinal ternary outcomes when some observed states are possibly misclassified. *Stat. Med.* **35** 1549–1557. MR3513468 https://doi.org/10.1002/sim.6861

BESSER, L., KUKULL, W., KNOPMAN, D. S., CHUI, H., GALASKO, D., WEINTRAUB, S. et al. (2018). Version 3 of the National Alzheimer's Coordinating Center's uniform data set. *Alzheimer Dis. Assoc. Disord.* **32** 351.

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

CAI, Z., ZENG, D., MARDER, K. S., HONIG, L. S. and WANG, Y. (2026). Supplement to "Dynamic Classification of Latent Disease Progression with Auxiliary Surrogate Labels." https://doi.org/10.1214/26-AOAS2150SUPP

COLLINS, L. M. and LANZA, S. T. (2009). *Latent Class and Latent Transition Analysis*: *With Applications in the Social*, *Behavioral*, *and Health Sciences*. Wiley, New York.

DERUITER, S. L., LANGROCK, R., SKIRBUTAS, T., GOLDBOGEN, J. A., CALAMBOKIDIS, J., FRIEDLAENDER, A. S. and SOUTHALL, B. L. (2017). A multivariate mixed hidden Markov model for blue whale be-

Key words and phrases. Disease progression, latent states, hidden Markov model, generative-discriminative model, Alzheimer's disease.

haviour and responses to sound exposure. *Ann. Appl. Stat.* **11** 362–392. MR3634328 https://doi.org/10.1214/16-AOAS1008

GUERREIRO, R., ESCOTT-PRICE, V., DARWENT, L., PARKKINEN, L., ANSORGE, O., HERNANDEZ, D. G. et al. (2016). Genome-wide analysis of genetic correlation in dementia with Lewy bodies, Parkinson's and Alzheimer's diseases. *Neurobiol. Aging* **38** 214–e7.

GUO, X., BOURGEOIS, F. T. and CAI, T. (2024). Quantifying proportion of treatment effect by surrogate endpoint under heterogeneity. *Stat. Methods Med. Res.* **33** 1152–1162. MR4792211 https://doi.org/10.1177/09622802241247719

HU, J. and SZYMCZAK, S. (2023). A review on longitudinal data analysis with random forest. *Brief. Bioinform.* **24** bbad002.

JACKSON, C. H. and SHARPLES, L. D. (2002). Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. *Stat. Med.* **21** 113–128.

JACKSON, C. H., SHARPLES, L. D., THOMPSON, S. G., DUFFY, S. W. and COUTO, E. (2003). Multistate Markov models for disease progression with classification error. *Statistician* **52** 193–209. MR1977260 https://doi.org/10.1111/1467-9884.00351

JAIN, L., KHRESTIAN, M., FORMICA, S., TUASON, E. D., PILLAI, J. A., RAO, S. et al. (2024). ATN cerebrospinal fluid biomarkers in dementia with Lewy bodies: Initial results from the United States dementia with Lewy bodies consortium. *Alzheimer's Dement.* **20** 549–562.

LIU, H., SONG, X., TANG, Y. and ZHANG, B. (2021). Bayesian quantile nonhomogeneous hidden Markov models. *Stat. Methods Med. Res.* **30** 112–128. MR4216850 https://doi.org/10.1177/0962280220942802

MARTINO, A., GUATTERI, G. and PAGANONI, A. M. (2020). Multivariate hidden Markov models for disease progression. *Stat. Anal. Data Min.* **13** 499–507. MR4176152 https://doi.org/10.1002/sam.11479

MARUOTTI, A. (2011). Mixed hidden Markov models for longitudinal data: An overview. *Int. Stat. Rev.* **79** 427–454.

MCDONNELL, E. I. (2021). Dynamic Graphical Models and Curve Registration for High-dimensional Time Course Data. PhD thesis, Columbia Univ.

MCKEITH, I. G., FERMAN, T. J., THOMAS, A. J., BLANC, F., BOEVE, B. F., FUJISHIRO, H. et al. (2020). Research criteria for the diagnosis of prodromal dementia with Lewy bodies. *Neurology* **94** 743–755.

MUTHÉN, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In *The Sage Handbook of Quantitative Methodology for the Social Sciences*, *Vol.* 345 106–109.

PRENTICE, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Stat. Med.* **8** 431–440.

PROUST-LIMA, C., SÉNE, M., TAYLOR, J. M. G. and JACQMIN-GADDA, H. (2014). Joint latent class models for longitudinal and time-to-event data: A review. *Stat. Methods Med. Res.* **23** 74–90. MR3190688 https://doi.org/10.1177/0962280212445839

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

RUBIN, M. L., CHAN, W., YAMAL, J.-M. and ROBERTSON, C. S. (2017). A joint logistic regression and covariate-adjusted continuous-time Markov chain model. *Stat. Med.* **36** 4570–4582. MR3731240 https://doi.org/10.1002/sim.7387

SALMON, D. P., GALASKO, D., HANSEN, L. A., MASLIAH, E., BUTTERS, N., THAL, L. J. and KATZMAN, R. (1996). Neuropsychological deficits associated with diffuse Lewy body disease. *Brain Cogn.* **31** 148–165.

SCOTT, G. D., ARNOLD, M. R., BEACH, T. G., GIBBONS, C. H., KANTHASAMY, A. G., LEBOVITZ, R. M. et al. (2022). Fluid and tissue biomarkers of Lewy body dementia: Report of an LBDA symposium. *Front. Neurol.* **12** 805135.

SHEN, J. and HE, X. (2015). Inference for subgroup analysis with a structured logistic-normal mixture model. *J. Amer. Statist. Assoc.* **110** 303–312. MR3338504 https://doi.org/10.1080/01621459.2014.894763

SONG, X., XIA, Y. and ZHU, H. (2017). Hidden Markov latent variable models with multivariate longitudinal data. *Biometrics* **73** 313–323. MR3632377 https://doi.org/10.1111/biom.12536

STUTE, W. (1986). On almost sure convergence of conditional empirical distribution functions. *Ann. Probab.* **14** 891–901. MR0841591

UNIV. WASHINGTON (2015). NACC UDS Researcher's Data Dictionary, version 3.0 ed. The National Alzheimer's Coordinating Center.

WANG, X., PARAST, L., TIAN, L. and CAI, T. (2020). Model-free approach to quantifying the proportion of treatment effect explained by a surrogate marker. *Biometrika* **107** 107–122. MR4064143 https://doi.org/10.1093/biomet/asz065

WANG, Y., CHEN, H., ZENG, D., MAURO, C., DUAN, N. and SHEAR, M. K. (2013). Auxiliary marker-assisted classification in the absence of class identifiers. *J. Amer. Statist. Assoc.* **108** 553–565. MR3174641 https://doi.org/10.1080/01621459.2013.775949

WEINER, M. F., RISSER, R. C., MUNRO CULLUM, C., HONIG, L., WHITE, C., SPECIALE, S. et al. (1996). Alzheimer's disease and its Lewy body variant: A clinical analysis of postmortem verified cases. *Amer. J. Psychiatry* **153** 1269–1273.

YOSHIZAWA, H., VONSATTEL, J. P. G. and HONIG, L. S. (2013). Early neuropsychological discriminants for Lewy body disease: An autopsy series. *J. Neurol. Neurosurg. Psychiatry* **84** 1326–1330.

ZHOU, J., SONG, X. and SUN, L. (2020). Continuous time hidden Markov model for longitudinal data. *J. Multivariate Anal.* **179** 104646,16. MR4110273 https://doi.org/10.1016/j.jmva.2020.104646

ZOU, Y., LIN, Y. and SONG, X. (2024). Bayesian heterogeneous hidden Markov models with an unknown number of states. *J. Comput. Graph. Statist.* **33** 15–24. MR4713939 https://doi.org/10.1080/10618600.2023.2231055

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 663–682
https://doi.org/10.1214/25-AOAS2109
© Institute of Mathematical Statistics, 2026

MOVING TOWARDS AUTOMATED INTERSTELLAR BOUNDARY EXPLORER DATA SELECTION WITH LOTUS

BY MADELINE A. STRICKLIN[1,a], LAUREN J. BEESLEY[1,b], BRIAN P. WEAVER[1,c], KELLY R. MORAN[1,d], DAVE OSTHUS[1,e], PAUL H. JANZEN[2,f], GRANT DAVID MEADORS[3,g] AND DANIEL B. REISENFELD[4,h]

[1]*Statistical Sciences Group, Los Alamos National Laboratory,* [a]*mstricklin@lanl.gov,* [b]*lvandervort@lanl.gov,* [c]*theguz@lanl.gov,* [d]*krmoran@lanl.gov,* [e]*dosthus@lanl.gov*

[2]*Department of Physics and Astronomy, University of Montana,* [f]*paul.janzen@mso.umt.edu*

[3]*Space Remote Sensing and Data Science Group, Los Alamos National Laboratory,* [g]*gdmeadors@lanl.gov*

[4]*Space Science and Applications Group, Los Alamos National Laboratory,* [h]*dreisenfeld@lanl.gov*

The Interstellar Boundary Explorer (IBEX) satellite collects data on energetic neutral atoms (ENAs) that provide insight into the heliosphere, the region surrounding our solar system and separating it from interstellar space. IBEX collects information on these particles and on extraneous "background" particles. While IBEX records how and when the different particles are observed, it does not distinguish between heliospheric ENA particles and incidental background particles. To address this issue, all IBEX data has historically been manually labeled as "good" ENA data, or "bad" background data. This manual culling process is incredibly time-intensive and contingent on subjective, manually-induced decision thresholds. In this paper, we develop a three-stage automated culling process, called LOTUS, that uses random forests to expedite and standardize the labeling process. In Stage 1, LOTUS uses random forests to obtain probabilities of observing true ENA particles on a per-observation basis. In Stage 2, LOTUS aggregates these probabilities to obtain predictions within small windows of time. In Stage 3, LOTUS refines these predictions. We compare the labels generated by LOTUS to those manually generated by the subject matter expert. We use various metrics to demonstrate that LOTUS is a useful automated process for supplementing and standardizing the manual culling process.

REFERENCES

ACTON, C., BACHMAN, N., SEMENOV, B. and WRIGHT, E. (2018). A look towards the future in the handling of space science mission geometry. *Planet. Space Sci.* **150** 9–12. https://doi.org/10.1016/j.pss.2017.02.013

ACTON, C. H. (1996). Ancillary data services of NASA's navigation and ancillary information facility. *Planet. Space Sci.* **44** 65–70. https://doi.org/10.1016/0032-0633(95)00107-7

ALLEGRINI, F., CREW, G., DEMKEE, D., FUNSTEN, H., MCCOMAS, D., RANDOL, B., RODRIGUEZ, B., SCHWADRON, N., VALEK, P. et al. (2008). The IBEX background monitor. *Space Sci. Rev.* **146** 105–115.

BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning*. *Information Science and Statistics*. Springer, New York.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series*: *Theory and Methods*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1093459 https://doi.org/10.1007/978-1-4419-0320-4

FREUND, R. J., WILSON, W. J. and MOHR, D. L. (2010). *Stat. Methods*. Academic Press, San Diego.

FUNSTEN, H., ALLEGRINI, F., BOCHSLER, P., DUNN, G., ELLIS, S., EVERETT, D., FAGAN, M., FUSELIER, S., GRANOFF, M. et al. (2009a). The interstellar boundary explorer high energy (IBEX-Hi) neutral atom imager. *Space Sci. Rev.* **146** 75–103.

FUNSTEN, H., ALLEGRINI, F., CREW, G., DEMAJISTRE, R., FRISCH, P., FUSELIER, S., GRUNTMAN, M., JANZEN, P., MCCOMAS, D. et al. (2009b). Structures and spectral variations of the outer heliosphere in IBEX energetic neutral atom maps. *Space Sci. Rev.* **326**. 2021. https://doi.org/10.1126/science.1180927

Key words and phrases. Physics-informed machine learning, classification, automation, pattern recognition, random forests, heliophysics.

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

LIN, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **1** 255–268. https://doi.org/10.2307/2532051

MASSEY, F. J. JR (1951). The Kolmogorov-Smirnov test for goodness of fit. *J. Amer. Statist. Assoc.* **46** 68–78.

MCCOMAS, D., ALLEGRINI, F., BOCHSLER, P., BZOWSKI, M., CHRISTIAN, E., CREW, G., DEMAJISTRE, R., FAHR, H., FICHTNER, H. et al. (2009). Global observations of the interstellar interaction from the Interstellar Boundary Explorer (IBEX). *Science* **326** 959–962.

MCCOMAS, D. J., CHRISTIAN, E. R., SCHWADRON, N. A., FOX, N., WESTLAKE, J., ALLEGRINI, F., BAKER, D. N., BIESECKER, D., BZOWSKI, M. et al. (2018). Interstellar Mapping and Acceleration Probe (IMAP): A new NASA mission. *Space Sci. Rev.* **214** 116. https://doi.org/10.1007/s11214-018-0550-1

MCNUTT, R. L., WIMMER-SCHWEINGRUBER, R. F., GRUNTMAN, M., KRIMIGIS, S. M., ROELOF, E. C., BRANDT, P. C., VERNON, S. R., PAUL, M. V., STOUGH, R. W. et al. (2022). Interstellar probe–destination: Universe! *Acta Astronaut.* **196** 13–28.

MILLER, R. G. JR. (2012). *Simultaneous Statistical Inference*, 2nd ed. Springer, New York.

NEMBRINI, S., KÖNIG, I. R. and WRIGHT, M. N. (2018). The revival of the Gini importance? *Bioinformatics* **34** 3711–3718.

OSTHUS, D., WEAVER, B. P., BEESLEY, L. J., MORAN, K. R., STRICKLIN, M. A., ZIRNSTEIN, E. J., JANZEN, P. H. and REISENFELD, D. B. (2023). Towards improved heliosphere sky map estimation with Theseus. *Technometrics* **66** 208–226. MR4740746 https://doi.org/10.1080/00401706.2023.2271017

R CORE TEAM (2021). *R*: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

REISENFELD, D. B., BZOWSKI, M., FUNSTEN, H. O., HEERIKHUISEN, J., JANZEN, P. H., KUBIAK, M. A., MCCOMAS, D. J., SCHWADRON, N. A., SOKÓŁ, J. M. et al. (2021). A three-dimensional map of the heliosphere from IBEX. *Astrophys. J.*, *Suppl. Ser.* **254**. https://doi.org/10.3847/1538-4365/abf658

STRICKLIN, M. A., BEESLEY, L. J., WEAVER, B. P., MORAN, K. R., OSTHUS, D., JANZEN, P. H., MEADORS, G. D. and REISENFELD, D. B. (2026). Supplement to “Moving towards automated interstellar boundary explorer data selection with LOTUS.” https://doi.org/10.1214/25-AOAS2109SUPP

WRIGHT, M. N. and ZIEGLER, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77** 1–17. https://doi.org/10.18637/jss.v077.i01

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 683–701
https://doi.org/10.1214/25-AOAS2110

FUNCTIONAL MIXTURE REGRESSION CONTROL CHART

BY CHRISTIAN CAPEZZA[1,a], FABIO CENTOFANTI[1,2,e], DAVIDE FORCINA[1,b], ANTONIO LEPORE[1,c] AND BIAGIO PALUMBO[1,d]

[1]*Department of Industrial Engineering, University of Naples Federico II,* [a]*christian.capezza@unina.it,* [b]*davide.forcina@unina.it,* [c]*antonio.lepore@unina.it,* [d]*biagio.palumbo@unina.it*

[2]*Section of Statistics and Data Science, Department of Mathematics, KU Leuven,* [e]*fabio.centofanti@kuleuven.be*

Industrial applications often exhibit multiple in-control patterns due to varying operating conditions, which makes a single functional linear model (FLM) inadequate to capture the complexity of the true relationship between a functional quality characteristic and covariates, which gives rise to the multimode profile monitoring problem. This issue is clearly illustrated in the resistance spot welding (RSW) process in the automotive industry, where different operating conditions lead to multiple in-control states. In these states, factors such as electrode tip wear and dressing can influence the functional quality characteristic differently, resulting in distinct FLMs across subpopulations. To address this problem, this article introduces the functional mixture regression control chart (FMRCC) to monitor functional quality characteristics with multiple in-control patterns and covariate information, modeled using a mixture of FLMs. A monitoring strategy based on the likelihood ratio test is proposed to monitor any deviation from the estimated in-control heterogeneous population. An extensive Monte Carlo simulation study is performed to compare the FMRCC with competing monitoring schemes that have already appeared in the literature, and a case study in the monitoring of an RSW process in the automotive industry, which motivated this research, illustrates its practical applicability.

REFERENCES

CAPEZZA, C., CAPIZZI, G., CENTOFANTI, F., LEPORE, A. and PALUMBO, B. (2024a). An adaptive multivariate functional EWMA control chart. *J. Qual. Technol.* **57** 1–15. https://doi.org/10.1080/00224065.2024.2383674

CAPEZZA, C., CENTOFANTI, F., FORCINA, D., LEPORE, A. and PALUMBO, B. (2026a). Supplement A to “Functional mixture regression control chart.” https://doi.org/10.1214/25-AOAS2110SUPPA

CAPEZZA, C., CENTOFANTI, F., FORCINA, D., LEPORE, A. and PALUMBO, B. (2026b). Supplement B to “Functional mixture regression control chart.” https://doi.org/10.1214/25-AOAS2110SUPPB

CAPEZZA, C., CENTOFANTI, F., LEPORE, A., MENAFOGLIO, A., PALUMBO, B. and VANTINI, S. (2023). funcharts: Control charts for multivariate functional data in R. *J. Qual. Technol.* **55** 566–583. https://doi.org/10.1080/00224065.2023.2219012

CAPEZZA, C., CENTOFANTI, F., LEPORE, A. and PALUMBO, B. (2021). Functional clustering methods for resistance spot welding process data in the automotive industry. *Appl. Stoch. Models Bus. Ind.* **37** 908–925. MR4323921 https://doi.org/10.1002/asmb.2648

CAPEZZA, C., CENTOFANTI, F., LEPORE, A. and PALUMBO, B. (2024b). Robust multivariate functional control chart. *Technometrics* **66** 531–547. MR4819271 https://doi.org/10.1080/00401706.2024.2327346

CAPEZZA, C., LEPORE, A. and PAYNABAR, K. (2025). Stream-Based Active Learning for Process Monitoring. *Technometrics*. https://doi.org/10.1080/00401706.2025.2561744

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

CENTOFANTI, F., KULAHCI, M., LEPORE, A. and SPOONER, M. P. (2025). Real-time monitoring of functional data. *J. Qual. Technol.* **57** 135–152. https://doi.org/10.1080/00224065.2024.2430978

CENTOFANTI, F., LEPORE, A., MENAFOGLIO, A., PALUMBO, B. and VANTINI, S. (2021b). Functional regression control chart. *Technometrics* **63** 281–294. MR4296897 https://doi.org/10.1080/00401706.2020.1753581

Key words and phrases. Functional data analysis, profile monitoring, statistical process control, functional mixture regression, multiple functional linear models.

CENTOFANTI, F., LEPORE, A. and PALUMBO, B. (2024). Sparse and smooth functional data clustering. *Statist. Papers* **65** 795–825. MR4721179 https://doi.org/10.1007/s00362-023-01408-1

CENTOFANTI, F., LEPORE, A. and PALUMBO, B. (2025). An adaptive multivariate functional control chart. *Technometrics*. **67** 603–616. https://doi.org/10.1080/00401706.2025.2491369

CHIOU, J.-M., CHEN, Y.-T. and YANG, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statist. Sinica* **24** 1571–1596. MR3308652

CHOU, S.-H., CHANG, S. I. and TSAI, T.-R. (2014). On monitoring of multiple non-linear profiles. *Int. J. Prod. Res.* **52** 3209–3224. https://doi.org/10.1080/00207543.2013.867088

CIARLEGLIO, A. and OGDEN, R. T. (2016). Wavelet-based scalar-on-function finite mixture regression models. *Comput. Statist. Data Anal.* **93** 86–96. MR3406197 https://doi.org/10.1016/j.csda.2014.11.017

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, Methodol.* **39** 1–38. With discussion. MR0501537

DESARBO, W. S. and CRON, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *J. Classification* **5** 249–282. MR0971156 https://doi.org/10.1007/BF01897167

DEVIJVER, E. (2017). Model-based regression clustering for high-dimensional data: Application to functional data. *Adv. Data Anal. Classif.* **11** 243–279. MR3656031 https://doi.org/10.1007/s11634-016-0242-1

DICKINSON, D., FRANKLIN, J., STANYA, A. et al. (1980). Characterization of spot welding behavior by dynamic electrical parameter monitoring. *Weld. J.* **59** 170.

EFRON, B. and TIBSHIRANI, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.* **1** 54–77. With a comment by J. A. Hartigan and a rejoinder by the authors. MR0833275

FRALEY, C. and RAFTERY, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Statist. Assoc.* **97** 611–631. MR1951635 https://doi.org/10.1198/016214502760047131

GRASSO, M., COLOSIMO, B. M. and TSUNG, F. (2017). A phase I multi-modelling approach for profile monitoring of signal data. *Int. J. Prod. Res.* **55** 4354–4377. https://doi.org/10.1080/00207543.2016.1251626

HAPP, C. and GREVEN, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J. Amer. Statist. Assoc.* **113** 649–659. MR3832216 https://doi.org/10.1080/01621459.2016.1273115

IANNONE, F., AMBROSINO, F., BRACCO, G., DE ROSA, M., FUNEL, A., GUARNIERI, G., MIGLIORI, S., PALOMBI, F., PONTI, G. et al. (2019). CRESCO ENEA HPC clusters: A working example of a multifabric GPFS spectrum scale layout. In 2019 *International Conference on High Performance Computing Simulation (HPCS)* 1051–1052.

JACQUES, J. and PREDA, C. (2013). Funclust: A curves clustering method using functional random variables density approximation. *Neurocomputing* **112** 164–171. https://doi.org/10.1016/j.neucom.2012.11.042

JONES, C. L., ABDEL-SALAM, A.-S. G. and MAYS, D. (2021). Practitioners guide on parametric, nonparametric, and semiparametric profile monitoring. *Qual. Reliab. Eng. Int.* **37** 857–881. https://doi.org/10.1002/qre.2770

JONES, P. and MCLACHLAN, G. J. (1992). Fitting finite mixture models in a regression context. *Aust. J. Stat.* **34** 233–240. https://doi.org/10.1111/j.1467-842X.1992.tb01356.x

KOKOSZKA, P. and REIMHERR, M. (2017). *Introduction to Functional Data Analysis*. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3793167

KRUGER, U. and XIE, L. (2012). *Statistical Monitoring of Complex Multivariate Processes: with Applications in Industrial Process Control*. *Statistics in Practice*. Wiley, Chichester. MR3024928 https://doi.org/10.1002/9780470517253.scard

MALEKI, M. R., AMIRI, A. and CASTAGLIOLA, P. (2018). An overview on recent profile monitoring papers (2008–2018) based on conceptual classification scheme. *Comput. Ind. Eng.* **126** 705–728. https://doi.org/10.1016/j.cie.2018.10.008

MANDEL, B. (1969). The regression control chart. *J. Qual. Technol.* **1** 1–9. https://doi.org/10.1080/00224065.1969.11980341

MANLADAN, S., YUSOF, F., RAMESH, S., FADZIL, M., LUO, Z. and AO, S. (2017). A review on resistance spot welding of aluminum alloys. *Int. J. Adv. Manuf. Technol.* **90** 605–634.

MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models*. *Wiley Series in Probability and Statistics: Applied Probability and Statistics*. Wiley Interscience, New York. MR1789474 https://doi.org/10.1002/0471721182

MCLACHLAN, G. J. and KRISHNAN, T. (2008). *The EM Algorithm and Extensions*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2392878 https://doi.org/10.1002/9780470191613

NOOROSSANA, R., SAGHAEI, A. and AMIRI, A. (2011). *Statistical Analysis of Profile Monitoring*. Wiley, New York.

PARK, C. and SHRIVASTAVA, A. K. (2014). Multimode geometric-profile monitoring with correlated image data and its application to nanoparticle self-assembly processes. *J. Qual. Technol.* **46** 216–233. https://doi.org/10.1080/00224065.2014.11917966

QIU, P. (2013). *Introduction to Statistical Process Control*. CRC press, Boca Raton, FL.

R CORE TEAM (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

RAMAKER, H.-J., VAN SPRANG, E. N., WESTERHUIS, J. A. and SMILDE, A. K. (2004). The effect of the size of the training set and number of principal components on the false alarm rate in statistical process monitoring. *Chemom*. *Intell*. *Lab*. *Syst*. **73** 181–187. https://doi.org/10.1016/j.chemolab.2003.12.015

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

SAIN, S. R., GRAY, H., WOODWARD, W. A. and FISK, M. D. (1999). Outlier detection from a mixture distribution when training data are unlabeled. *Bull*. *Seismol*. *Soc*. *Amer*. **89** 294–304. https://doi.org/10.1785/BSSA0890010294

SCRUCCA, L., FOP, M., MURPHY, T. B. and RAFTERY, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R J*. **8** 289.

VALAEE-TALE, M., SHEIKHI, M., MAZAHERI, Y., GHAINI, F. M. and USEFIFAR, G. R. (2020). Criterion for predicting expulsion in resistance spot welding of steel sheets. *J*. *Mater*. *Process*. *Technol*. **275** 116329. https://doi.org/10.1016/j.jmatprotec.2019.116329

WANG, K., LI, J. and TSUNG, F. (2018). Registration-free monitoring of multimode near-circular shape profiles. *Qual*. *Reliab*. *Eng*. *Int*. **34** 529–542. https://doi.org/10.1002/qre.2270

WANG, K., NG, S.-K. and MCLACHLAN, G. J. (2009). Multivariate skew t mixture models: Applications to fluorescence-activated cell sorting data. In 2009 *Digital Image Computing*: *Techniques and Applications* 526–531. IEEE Press, New York.

WANG, S., HUANG, M., WU, X. and YAO, W. (2016). Mixture of functional linear models and its application to CO_2-GDP functional data. *Comput*. *Statist*. *Data Anal*. **97** 1–15. MR3447032 https://doi.org/10.1016/j.csda.2015.11.008

WANG, S., WOODWARD, W. A., GRAY, H. L., WIECHECKI, S. and SAIN, S. R. (1997). A new test for outlier detection from a multivariate mixture distribution. *J*. *Comput*. *Graph*. *Statist*. **6** 285–299. MR1466869 https://doi.org/10.2307/1390734

WOODALL, W. H., SPITZNER, D. J., MONTGOMERY, D. C. and GUPTA, S. (2004). Using control charts to monitor process and product quality profiles. *J*. *Qual*. *Technol*. **36** 309–320. https://doi.org/10.1080/00224065.2004.11980276

WU, C.-F. J. (1983). On the convergence properties of the EM algorithm. *Ann*. *Statist*. **11** 95–103. MR0684867 https://doi.org/10.1214/aos/1176346060

YAO, F., FU, Y. and LEE, T. C. (2011). Functional mixture regression. *Biostatistics* **12** 341–353. https://doi.org/10.1093/biostatistics/kxq067

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

ZHAO, Y., TODD OGDEN, R. and REISS, P. T. (2012). Wavelet-based LASSO in functional linear regression. *J*. *Comput*. *Graph*. *Statist*. **21** 600–617. MR2970910 https://doi.org/10.1080/10618600.2012.679241

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 702–723
https://doi.org/10.1214/25-AOAS2125

NEURAL POSTERIOR ESTIMATION WITH AUTOREGRESSIVE TILING FOR DETECTING OBJECTS IN ASTRONOMICAL IMAGES

BY JEFFREY REGIER[a]

Department of Statistics, University of Michigan, [a]*regier@umich.edu*

Upcoming astronomical surveys will produce petabytes of high-resolution images of the night sky, providing information about billions of stars and galaxies. Detecting and characterizing the astronomical objects in these images is a fundamental task in astronomy—and a challenging one, as most of these objects are faint and many visually overlap with other objects. We propose an amortized variational inference procedure to solve this instance of small-object detection. Our key innovation is a family of spatially autoregressive variational distributions that partition and order the latent space according to a K-color checkerboard pattern. By construction, the conditional independencies of this variational family mirror those of the posterior distribution. We fit the variational distribution, which is parameterized by a convolutional neural network, using neural posterior estimation (NPE) to minimize an expectation of the forward KL divergence. Using images from the Sloan Digital Sky Survey, the proposed method achieves state-of-the-art performance. We further demonstrate that the proposed autoregressive structure greatly improves posterior calibration.

REFERENCES

AMBROGIONI, L., GÜÇLÜ, U., BEREZUTSKAYA, J., BORNE, E., GÜÇLÜTÜRK, Y., HINNE, M., MARIS, E. and GERVEN, M. (2019). Forward amortized inference for likelihood-free variational marginalization. In *International Conference on Artificial Intelligence and Statistics*.

BARBER, D. and AGAKOV, F. V. (2004). The IM algorithm: A variational approach to information maximization. In *Advances in Neural Information Processing Systems*.

BLANTON, M. R., HOGG, D. W., BAHCALL, N. A., BALDRY, I. K., BRINKMANN, J., CSABAI, I., EISENSTEIN, D., FUKUGITA, M., GUNN, J. E. et al. (2003). The broadband optical properties of galaxies with redshifts $0.02 < Z < 0.22$. *Astrophys. J.* **594** 186.

BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 https://doi.org/10.1080/01621459.2017.1285773

BORNSCHEIN, J. and BENGIO, Y. (2015). Reweighted wake-sleep. In *International Conference on Learning Representations*.

BOSCH, J., ALSAYYAD, Y., ARMSTRONG, R., BELLM, E., CHIANG, H.-F., EGGL, S., FINDEISEN, K., FISHER-LEVINE, M., GUY, L. P. et al. (2018a). An overview of the LSST image processing pipelines. arXiv preprint. Available at arXiv:1812.03248.

BOSCH, J., ARMSTRONG, R., BICKERTON, S. et al. (2018b). The hyper suprime-cam software pipeline. *Publ. Astron. Soc. Jpn.* **70** 1–39.

BREWER, B. J., FOREMAN-MACKEY, D. and HOGG, D. W. (2013). Probabilistic catalogs for crowded stellar fields. *Astron. J.* **146** 7–15.

BUCHANAN, J. J., SCHNEIDER, M. D., PRUETT, K. and ARMSTRONG, R. E. (2023). Markov chain Monte Carlo for Bayesian parametric galaxy modeling in LSST. arXiv preprint. Available at arXiv:2309.10321.

CONSELICE, C. J. (2014). The evolution of galaxy structure over cosmic time. *Annu. Rev. Astron. Astrophys.* **52** 291–337.

DEVLIN, J., CHANG, M.-W., LEE, K. and TOUTANOVA, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the* 2019 *Conference of the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies* 4171–4186.

FEDER, R. M., PORTILLO, S. K., DAYLAN, T. and FINKBEINER, D. (2020). Multiband probabilistic cataloging: A joint fitting approach to point-source detection and deblending. *Astron. J.* **159** 163.

Key words and phrases. Variational inference, amortized inference, simulation-based inference, likelihood-free inference, object detection, computer vision.

HANSEN, D., MENDOZA, I., LIU, R., PANG, Z., ZHAO, Z., AVESTRUZ, C. and REGIER, J. (2022). Scalable Bayesian inference for detection and deblending in astronomical images. ICML Workshop on Machine Learning for Astrophysics.

LIU, R., MCAULIFFE, J. D., REGIER, J. and COLLABORATION, T. L. D. E. S. (2023). Variational inference for deblending crowded starfields. *J. Mach. Learn. Res.* **24** 1–36. MR4633568

LUPTON, R., GUNN, J. E., IVEZIC, Z., KNAPP, G. R., KENT, S. and YASUDA, N. (2001). The SDSS imaging pipelines. arXiv preprint. Available at arXiv:astro-ph/0101420.

MAALØE, L., SØNDERBY, C. K., SØNDERBY, S. K. and WINTHER, O. (2016). Auxiliary deep generative models. In *International Conference on Machine Learning* 1445–1453.

MALZ, A. and HOGG, D. (2022). How to obtain the redshift distribution from probabilistic redshift estimates. *Astrophys. J.* **928** 127.

MANDELBAUM, R. (2018). Weak lensing for precision cosmology. *Annu. Rev. Astron. Astrophys.* **56** 393–433.

MELCHIOR, P., JOSEPH, R., SANCHEZ, J., MACCRANN, N. and GRUEN, D. (2021). The challenge of blending in large sky surveys. *Nat. Rev. Phys.* **3** 712–718.

MODRÁK, M., MOON, A. H., KIM, S., BÜRKNER, P., HUURRE, N., FALTEJSKOVÁ, K., GELMAN, A. and VEHTARI, A. (2025). Simulation-based calibration checking for Bayesian computation: The choice of test quantities shapes sensitivity. *Bayesian Anal.* **20** 461–488. MR4905015 https://doi.org/10.1214/23-ba1404

MORGANSON, E., GRUENDL, R. A., MENANTEAU, F. et al. (2018). The dark energy survey image processing pipeline. *Publ. Astron. Soc. Pac.* **130** 074501.

RUBIN OBSERVATORY (2025). Key numbers. https://rubinobservatory.org/for-scientists/rubin-101/key-numbers. Accessed: 2025-08-31.

PAPAMAKARIOS, G. and MURRAY, I. (2016). Fast ϵ-free inference of simulation models with Bayesian conditional density estimation. In *Neural Information Processing Systems*.

PATEL, A., ZHANG, T., AVESTRUZ, C., REGIER, J. and THE LSST DARK ENERGY SCIENCE COLLABORATION (2025). Neural posterior estimation for cataloging astronomical images with spatially varying backgrounds and point spread functions. *Astron. J.* **170** 155.

PATEL, Y. and REGIER, J. (2022). Scalable Bayesian inference for detecting strong gravitational lensing systems. In *NeurIPS Workshop on Machine Learning and the Physical Sciences*.

PORTILLO, S. K., LEE, B. C., DAYLAN, T. and FINKBEINER, D. P. (2017). Improved point-source detection in crowded fields using probabilistic cataloging. *Astron. J.* **154** 132.

REGIER, J. (2026). Supplement to "Neural posterior estimation with autoregressive tiling for detecting objects in astronomical images." https://doi.org/10.1214/25-AOAS2125SUPPA, https://doi.org/10.1214/25-AOAS2125SUPPB

REGIER, J., MILLER, A. C., SCHLEGEL, D., ADAMS, R. P., MCAULIFFE, J. D. and PRABHAT (2019). Approximate inference for constructing astronomical catalogs from images. *Ann. Appl. Stat.* **13** 1884–1926. MR4019161 https://doi.org/10.1214/19-AOAS1258

ROMELLI, E., FRAILIS, M., GALEOTTA, S., TAVAGNACCO, D., MAINO, D., VUERLI, C., MAGGIO, G. and TAFFONI, G. (2019). Euclidizing external tools: An example from SDC-IT on how to handle software and humanware. *Astron. Data Anal. Softw. Syst. XXVII* **523** 199.

ROWE, B., JARVIS, M., MANDELBAUM, R., BERNSTEIN, G. et al. (2015). GalSim: The modular galaxy image simulation toolkit. *Astron. Comput.* **10** 121–150.

RYKOFF, E., ROZO, E., BUSHA, M., CUNHA, C., FINOGUENOV, A., EVRARD, A., HAO, J., KOESTER, B., LEAUTHAUD, A. et al. (2014). redMaPPer. I. Algorithm and SDSS DR8 catalog. *Astrophys. J.* **785** 104.

SANCHEZ, J., MENDOZA, I., KIRKBY, D. P. and BURCHAT, P. R. (2021). Effects of overlapping sources on cosmic shear estimation: Statistical sensitivity and pixel-noise bias. *J. Cosmol. Astropart. Phys.* **2021** 43.

SCHAFER, C. M. (2015). A framework for statistical inference in astrophysics. *Annu. Rev. Stat. Appl.* **2** 141–162.

SCHNEIDER, M., NG, K., DAWSON, W., MARSHALL, P., MEYERS, J. and BARD, D. (2017). Probabilistic cosmological mass mapping from weak lensing shear. *Astrophys. J.* **839** 25.

STETSON, P. B. (1987). DAOPHOT: A computer program for crowded-field stellar photometry. *Astron. Soc. Pac.* **99** 191–222.

STRAUSS, M. A., WEINBERG, D. H., LUPTON, R. H., NARAYANAN, V. K., ANNIS, J., BERNARDI, M., BLANTON, M., BURLES, S., CONNOLLY, A. et al. (2002). Spectroscopic target selection in the sloan digital sky survey: The main galaxy sample. *Astron. J.* **124** 1810.

WANG, C. and SENNRICH, R. (2020). On exposure bias, hallucination and domain shift in neural machine translation. In *Proceedings of the* 58*th Annual Meeting of the Association for Computational Linguistics* 3544–3552. Association for Computational Linguistics.

WEBB, S., GOLINSKI, A., ZINKOV, R., RAINFORTH, T., TEH, Y. W., WOOD, F. et al. (2018). Faithful inversion of generative models for effective amortized inference. In *Neural Information Processing Systems*.

YAO, Y., VEHTARI, A., SIMPSON, D. and GELMAN, A. (2018). Yes, but did it work?: Evaluating variational inference. In *International Conference on Machine Learning*.

ZHANG, C., BÜTEPAGE, J., KJELLSTRÖM, H. and MANDT, S. (2018). Advances in variational inference. *IEEE Trans. Pattern Anal. Mach. Intell.* **41** 2008–2026.

ZHANG, L., BLEI, D. and NAESSETH, C. A. (2023). Transport score climbing: Variational inference using forward KL and adaptive neural transport. *Trans. Mach. Learn. Res.*.

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 724–743
https://doi.org/10.1214/26-AOAS2151

DO LARGE LANGUAGE MODELS (REALLY) NEED STATISTICAL FOUNDATIONS?

BY WEIJIE SU[a]

Department of Statistics and Data Science, University of Pennsylvania, [a]*suw@wharton.upenn.edu*

Large language models (LLMs) represent a new paradigm for processing unstructured data, with applications across an unprecedented range of domains. In this paper we address, through two arguments, whether the development and application of LLMs would genuinely benefit from foundational contributions from the statistics discipline. First, we argue affirmatively, beginning with the observation that LLMs are inherently statistical models due to their profound data dependency and stochastic generation processes, where statistical insights are naturally essential for handling variability and uncertainty. Second, we argue that the persistent black-box nature of LLMs—stemming from their immense scale, architectural complexity, and development practices often prioritizing empirical performance over theoretical interpretability—renders closed-form or purely mechanistic analyses generally intractable, thereby necessitating statistical approaches due to their flexibility and often demonstrated effectiveness. To substantiate these arguments, the paper outlines several research areas—including alignment, watermarking, uncertainty quantification, evaluation, and data mixture optimization—where statistical methodologies are critically needed and are already beginning to make valuable contributions. We conclude with a discussion suggesting that statistical research concerning LLMs will likely form a diverse "mosaic" of specialized topics, rather than deriving from a single unifying theory, and highlight the importance of timely engagement by our statistics community in LLM research.

REFERENCES

AARONSON, S. (2023). Watermarking of large language models. Available at https://simons.berkeley.edu/talks/scott-aaronson-ut-austin-openai-2023-08-17.

ACHIAM, J., ADLER, S., AGARWAL, S., AHMAD, L., AKKAYA, I., ALEMAN, F. L., ALMEIDA, D., ALTENSCHMIDT, J., ALTMAN, S. et al. (2023). GPT-4 technical report. arXiv preprint. Available at arXiv:2303.08774.

ADLER, B., AGARWAL, N., AITHAL, A., ANH, D. H., BHATTACHARYA, P., BRUNDYN, A., CASPER, J., CATANZARO, B., CLAY, S. et al. (2024). Nemotron-4 340b technical report. arXiv preprint. Available at arXiv:2406.11704.

ALAA, A. and YU, B. (2024). Veridical data science for medical foundation models. arXiv preprint. Available at arXiv:2409.10580.

ANGELOPOULOS, A. N., BATES, S., FANNJIANG, C., JORDAN, M. I. and ZRNIC, T. (2023). Prediction-powered inference. *Science* **382** 669–674. MR4672910

BAI, X., WANG, A., SUCHOLUTSKY, I. and GRIFFITHS, T. L. (2025). Explicitly unbiased large language models still form biased associations. *Proc. Natl. Acad. Sci. USA* **122** e2416228122.

BENDER, E. M., GEBRU, T., MCMILLAN-MAJOR, A. and SHMITCHELL, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the* 2021 *ACM Conference on Fairness*, *Accountability*, *and Transparency* (*FAccT* '21) 610–623.

BOX, G. E. P. (1976). Science and statistics. *J. Amer. Statist. Assoc.* **71** 791–799. MR0431440

BRADLEY, R. A. and TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39** 324–345. MR0070925 https://doi.org/10.2307/2334029

BREIMAN, L. (2001). Statistical modeling: The two cultures. *Statist. Sci.* **16** 199–231. With comments and a rejoinder by the author. MR1874152 https://doi.org/10.1214/ss/1009213726

Key words and phrases. Large language models, statistical foundations, statistical models, black-box models.

BROWN, T., MANN, B., RYDER, N., SUBBIAH, M., KAPLAN, J. D., DHARIWAL, P., NEELAKANTAN, A., SHYAM, P., SASTRY, G. et al. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems* **33** 1877–1901.

BUBECK, S., CHANDRASEKARAN, V., ELDAN, R., GEHRKE, J., HORVITZ, E., KAMAR, E., LEE, P., LEE, Y. T., LI, Y. et al. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv preprint. Available at arXiv:2303.12712.

BUBECK, S. and SELLKE, M. (2021). A universal law of robustness via isoperimetry. *Adv. Neural Inf. Process. Syst.* **34** 28811–28822.

CAI, Y., LI, L. and ZHANG, L. (2025). A statistical hypothesis testing framework for data misappropriation detection in large language models. arXiv preprint. Available at arXiv:2501.02441.

CANDÈS, E. and SABATTI, C. (2020). Discussion of the paper "Prediction, estimation, and attribution" by B. Efron [MR4107663]. *J. Amer. Statist. Assoc.* **115** 656–658. MR4107664 https://doi.org/10.1080/01621459.2020.1762618

CAO, Y., HE, Y., WU, D., CHEN, H.-Y., FAN, J. and LIU, H. (2025). Transformers simulate MLE for sequence generation in Bayesian networks. arXiv preprint. Available at arXiv:2501.02547.

CAO, Y. and YANG, J. (2015). Towards making systems forget with machine unlearning. In 2015 *IEEE Symposium on Security and Privacy* 463–480. IEEE.

CHAKRABORTY, S., QIU, J., YUAN, H., KOPPEL, A., HUANG, F., MANOCHA, D., BEDI, A. S. and WANG, M. (2024). MaxMin-RLHF: Alignment with diverse human preferences. In *International Conference on Machine Learning* 6116–6135. PMLR.

CHEN, C., BORGEAUD, S., IRVING, G., LESPIAU, J.-B., SIFRE, L. and JUMPER, J. (2023). Accelerating large language model decoding with speculative sampling. arXiv preprint. Available at arXiv:2302.01318.

CHERIAN, J. J., GIBBS, I. and CANDÈS, E. J. (2024). Large language model validity via enhanced conformal prediction methods. *Adv. Neural Inf. Process. Syst.* **37** 114812–114842.

CHIANG, T. (2023). ChatGPT is a blurry jpeg of the web. Available at https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web. Accessed: 2025-05-05.

COBBE, K., KOSARAJU, V., BAVARIAN, M., CHEN, M., JUN, H., KAISER, L., PLAPPERT, M., TWOREK, J., HILTON, J. et al. (2021). Training verifiers to solve math word problems. arXiv preprint. Available at arXiv:2110.14168.

COLLINS, K. M., JIANG, A. Q., FRIEDER, S., WONG, L., ZILKA, M., BHATT, U., LUKASIEWICZ, T., WU, Y., TENENBAUM, J. B. et al. (2024). Evaluating language models for mathematics through interactions. *Proc. Natl. Acad. Sci. USA* **121** Paper No. e2318124121, 11. MR4771734

DAI, D., DENG, C., ZHAO, C., XU, R., GAO, H., CHEN, D., LI, J., ZENG, W., YU, X. et al. (2024). Deepseekmoe: Towards ultimate expert specialization in mixture-of-experts language models. In *Proceedings of the* 62*nd Annual Meeting of the Association for Computational Linguistics* (*Volume* 1: *Long Papers*) 1280–1297.

DENG, J., LI, T.-W., ZHANG, S., LIU, S., PAN, Y., HUANG, H., WANG, X., HU, P., ZHANG, X. et al. (2024). dattri: A library for efficient data attribution. *Adv. Neural Inf. Process. Syst.* **37** 136763–136781.

DEVLIN, J., CHANG, M.-W., LEE, K. and TOUTANOVA, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the* 2019 *Conference of the North American Chapter of the Association for Computational Linguistics* 4171–4186.

DEY, A. and DONOHO, D. (2024). Universality of the $\pi^2/6$ pathway in avoiding model collapse. arXiv preprint. Available at arXiv:2410.22812.

DIMA, A., FOULDS, J., PAN, S. and FELDMAN, P. (2025). You've changed: Detecting modification of black-box large language models. arXiv preprint. Available at arXiv:2504.12335.

DONOHO, D. (2017). 50 years of data science. *J. Comput. Graph. Statist.* **26** 745–766. MR3765335 https://doi.org/10.1080/10618600.2017.1384734

DONOHO, D. (2024). Data science at the singularity. *Harv. Data Sci. Rev.* **6**.

DOU, Z., KOTEKAL, S., XU, Z. and ZHOU, H. H. (2024). From optimal score matching to optimal sampling. arXiv preprint. Available at arXiv:2409.07032.

DUAN, H., YANG, Y. and TAM, K. Y. (2024). Do LLMs know about hallucination? An empirical investigation of LLM's hidden states. arXiv preprint. Available at arXiv:2402.09733.

DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*: *Third Theory of Cryptography Conference*, *TCC* 2006, *New York*, *NY*, *USA*, *March* 4–7, 2006. *Proceedings* 3. *Lecture Notes in Computer Science* **3876** 265–284. Springer, Berlin. MR2241676 https://doi.org/10.1007/11681878_14

EFRON, B. (2020). Prediction, estimation, and attribution. *J. Amer. Statist. Assoc.* **115** 636–655. MR4107663 https://doi.org/10.1080/01621459.2020.1762613

EFROS, A. (2023). Ai is (mostly) an experimental science. Work like a biologist rather than a mathematician. X (formerly Twitter). Available at https://x.com/turingbook/status/1679718448249864194.

ELDAN, R. and LI, Y. (2023). Tinystories: How small can language models be and still speak coherent English? arXiv preprint. Available at arXiv:2305.07759.

ELHAGE, N., NANDA, N., OLSSON, C., HENIGHAN, T., JOSEPH, N., MANN, B., ASKELL, A., BAI, Y., CHEN, A. et al. (2021). A mathematical framework for transformer circuits. *Transform. Circuits Thread* **1** 12.

FENG, Y., KWIATKOWSKI, A., ZHENG, K., KEMPE, J. and DUAN, Y. (2025). PILAF: Optimal human preference sampling for reward modeling. In *Forty-Second International Conference on Machine Learning*.

GAGE, P. (1994). A new algorithm for data compression. *C Users J.* **12** 23–38.

GAO, T., JIN, J., KE, Z. T. and MORYOUSSEF, G. (2026). A Comparison of DeepSeek and other LLMs. *Amer. Statist.* **80** 164–176. MR5033795 https://doi.org/10.1080/00031305.2025.2611010

GERSTGRASSER, M., SCHAEFFER, R., DEY, A., RAFAILOV, R., SLEIGHT, H., HUGHES, J., KORBAK, T., AGRAWAL, R., PAI, D. et al. (2024). Is model collapse inevitable? Breaking the curse of recursion by accumulating real and synthetic data. In *First Conference on Language Modeling*.

GRATTAFIORI, A., DUBEY, A., JAUHRI, A., PANDEY, A., KADIAN, A., AL-DAHLE, A., LETMAN, A., MATHUR, A., SCHELTEN, A. et al. (2024). The llama 3 herd of models. arXiv preprint. Available at arXiv:2407.21783.

GU, A. and DAO, T. (2024). Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*.

GUI, Y., JIN, Y. and REN, Z. (2024). Conformal alignment: Knowing when to trust foundation models with guarantees. In *Advances in Neural Information Processing Systems*.

GUO, D., YANG, D., ZHANG, H., SONG, J., ZHANG, R., XU, R., ZHU, Q., MA, S., WANG, P. et al. (2025). DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. arXiv preprint. Available at arXiv:2501.12948.

GUPTA, V., KOREN, T. and SINGER, Y. (2018). Shampoo: Preconditioned stochastic tensor optimization. In *International Conference on Machine Learning* 1842–1850. PMLR.

HE, B. and HOFMANN, T. (2024). Simplifying transformer blocks. In *ICLR* 2024.

HE, Y., CAO, Y., CHEN, H.-Y., WU, D., FAN, J. and LIU, H. (2025). Learning spectral methods by transformers. arXiv preprint. Available at arXiv:2501.01312.

HENDRYCKS, D., BURNS, C., BASART, S., ZOU, A., MAZEIKA, M., SONG, D. and STEINHARDT, J. (2021). Measuring massive multitask language understanding. In *International Conference on Learning Representations*.

HINTON, G. (2023). GPT-4 as humanity's butterfly. Twitter post. Available at https://x.com/geoffreyhinton/status/1635739459764322330.

HOFFMANN, J., BORGEAUD, S., MENSCH, A., BUCHATSKAYA, E., CAI, T., RUTHERFORD, E., DE LAS CASAS, D., HENDRICKS, L. A., WELBL, J. et al. (2022). Training compute-optimal large language models. In *Proceedings of the* 36*th International Conference on Neural Information Processing Systems* 30016–30030.

HU, Z. and HUANG, H. (2024). Inevitable trade-off between watermark strength and speculative sampling efficiency for language models. arXiv preprint. Available at arXiv:2410.20418.

HUANG, L., YU, W., MA, W., ZHONG, W., FENG, Z., WANG, H., CHEN, Q., PENG, W., FENG, X. et al. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Trans. Inf. Syst.* **43** 1–55.

JESSON, A., BELTRAN-VELEZ, N., CHU, Q., KARLEKAR, S., KOSSEN, J., GAL, Y., CUNNINGHAM, J. P. and BLEI, D. (2024). Estimating the hallucination rate of generative ai. *Adv. Neural Inf. Process. Syst.* **37** 31154–31201.

JI, W., YUAN, W., GETZEN, E., CHO, K., JORDAN, M. I., MEI, S., WESTON, J. E., SU, W. J., XU, J. et al. (2025). An overview of large language models for statisticians. arXiv preprint. Available at arXiv:2502.17814.

JONES, C. R. and BERGEN, B. K. (2025). Large language models pass the Turing test. arXiv preprint. Available at arXiv:2503.23674.

JORDAN, K., JIN, Y., BOZA, V., YOU, J., CESISTA, F., NEWHOUSE, L. and BERNSTEIN, J. (2024). Muon: an optimizer for hidden layers in neural networks. Available at https://kellerjordan.github.io/posts/muon/. Accessed: 2025-05-24.

KALAI, A. T., NACHUM, O., VEMPALA, S. S. and ZHANG, E. (2025). Why language models hallucinate. arXiv preprint. Available at arXiv:2509.04664.

KALAI, A. T. and VEMPALA, S. S. (2024). Calibrated language models must hallucinate. In *STOC'*24*—Proceedings of the* 56*th Annual ACM Symposium on Theory of Computing* 160–171. ACM, New York. MR4764802 https://doi.org/10.1145/3618260.3649777

KAPLAN, J., MCCANDLISH, S., HENIGHAN, T., BROWN, T. B., CHESS, B., CHILD, R., GRAY, S., RADFORD, A., WU, J. et al. (2020). Scaling laws for neural language models. arXiv preprint. Available at arXiv:2001.08361.

KHOMENKO, L. (2025). Too many AIs. Available at https://dev.to/leeaao/too-many-ais-24nb.

KINGMA, D. and BA, J. (2015). A method for stochastic optimization. In *International Conference on Learning Representations*.

KIRCHENBAUER, J., GEIPING, J., WEN, Y., KATZ, J., MIERS, I. and GOLDSTEIN, T. (2023). A watermark for large language models. In *International Conference on Machine Learning* 17061–17084. PMLR.

KOH, P. W. and LIANG, P. (2017). Understanding black-box predictions via influence functions. In *International Conference on Machine Learning* 1885–1894. PMLR.

KOTEK, H., DOCKUM, R. and SUN, D. (2023). Gender bias and stereotypes in large language models. In *Proceedings of the ACM Collective Intelligence Conference* 12–24.

KOTTKE, J. (2023). The octopus test for large language model AIs. Available at https://kottke.org/23/03/the-octopus-test-for-large-language-model-ais. Accessed: 2025-05-05.

KRIZHEVSKY, A., SUTSKEVER, I. and HINTON, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25** 1097–1105.

LAU, T. T.-K., LONG, Q. and SU, W. (2025). PolarGrad: a class of matrix-gradient optimizers from a unifying preconditioning perspective. arXiv preprint. Available at arXiv:2505.21799.

LEVIATHAN, Y., KALMAN, M. and MATIAS, Y. (2023). Fast inference from transformers via speculative decoding. In *International Conference on Machine Learning* 19274–19286. PMLR.

LI, X., RUAN, F., WANG, H., LONG, Q. and SU, W. J. (2025a). Robust detection of watermarks for large language models under human edits. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*

LI, X., RUAN, F., WANG, H., LONG, Q. and SU, W. J. (2025b). A statistical framework of watermarks for large language models: Pivot, detection efficiency and optimal rules. *Ann. Statist.* **53** 322–351. MR4865018 https://doi.org/10.1214/24-aos2468

LI, X., TRAMER, F., LIANG, P. and HASHIMOTO, T. (2022). Large language models can be strong differentially private learners. In *International Conference on Learning Representations*.

LI, X., XIN, J., LONG, Q. and SU, W. J. (2025c). Evaluating the unseen capabilities: How many theorems do LLMs know? arXiv preprint. Available at arXiv:2506.02058.

LI, Y., WEI, F., ZHANG, C. and EAGLE, H. Z. (2024). Speculative sampling requires rethinking feature uncertainty. In *Proceedings of the 41st International Conference on Machine Learning* 28935–28948.

LIN, C. W. and ZHU, W. (2025). Divergent llm adoption and heterogeneous convergence paths in research writing. arXiv preprint. Available at arXiv:2504.13629.

LIN, S., HILTON, J. and EVANS, O. (2022). TruthfulQA: Measuring how models mimic human falsehoods. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 3214–3252.

LIN, X., CAI, T., DONOHO, D., FU, H., KE, T., JIN, J., MENG, X.-L., QU, A., SHI, C. et al. (2025). Statistics and AI: A fireside conversation. *Harv. Data Sci. Rev.* **7**.

LIU, A., FENG, B., XUE, B., WANG, B., WU, B., LU, C., ZHAO, C., DENG, C., ZHANG, C. et al. (2024a). DeepSeek-v3 technical report. arXiv preprint. Available at arXiv:2412.19437.

LIU, K., LONG, Q., SHI, Z., SU, W. J. and XIAO, J. (2025a). Statistical impossibility and possibility of aligning LLMs with human preferences: from Condorcet paradox to Nash equilibrium. arXiv preprint. Available at arXiv:2503.10990.

LIU, Q., ZHENG, X., MUENNIGHOFF, N., ZENG, G., DOU, L., PANG, T., JIANG, J. and LIN, M. (2025b). Regmix: Data mixture as regression for language model pre-training. In *International Conference on Learning Representations*.

LIU, X., CHEN, T., DA, L., CHEN, C., LIN, Z. and WEI, H. (2025c). Uncertainty quantification and confidence calibration in large language models: A survey. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining v. 2* 6107–6117.

LIU, Y., YANG, K., QI, Z., LIU, X., YU, Y. and ZHAI, C. X. (2024b). Bias and volatility: A statistical framework for evaluating large language model's stereotypes and the associated generation inconsistency. *Adv. Neural Inf. Process. Syst.* **37** 110131–110155.

LOSHCHILOV, I. and HUTTER, F. (2019). Decoupled weight decay regularization. In *International Conference on Learning Representations*.

MADAAN, L., SINGH, A. K., SCHAEFFER, R., POULTON, A., KOYEJO, S., STENETORP, P., NARANG, S. and HUPKES, D. (2024). Quantifying variance in evaluation benchmarks. arXiv preprint. Available at arXiv:2406.10229.

MANNING, C. D. and SCHÜTZE, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA. MR1722790

MATTHEW, B. and TONER, H. (2024). The surprising power of next word prediction: Large language models explained, Part 1. Georgetown Univ. Available at https://cset.georgetown.edu/article/the-surprising-power-of-next-word-prediction-large-language-models-explained-part-1/. Accessed May 5, 2025. Explains the mechanism and notes limitations are discussed in subsequent parts.

MIKOLOV, T., SUTSKEVER, I., CHEN, K., CORRADO, G. S. and DEAN, J. (2013). Distributed representations of words and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* **26**.

MILLER, E. (2024). Adding error bars to evals: a statistical approach to language model evaluations. arXiv preprint. Available at arXiv:2411.00640.

MOHRI, C. and HASHIMOTO, T. (2024). Language models with conformal factuality guarantees. In *International Conference on Machine Learning* 36029–36047. PMLR.

MUENNIGHOFF, N., YANG, Z., SHI, W., LI, X. L., FEI-FEI, L., HAJISHIRZI, H., ZETTLEMOYER, L., LIANG, P. and CANDÈS, E. (2025). s1: Simple test-time scaling. In *Proceedings of the* 2025 *Conference on Empirical Methods in Natural Language Processing* 20286–20332.

MUKHERJEE, S. (2025). Adaptive Data Collection for Policy Evaluation, Multi-task Learning and LLM Alignment. Univ. Wisconsin-Madison.

NIE, S., ZHU, F., YOU, Z., ZHANG, X., OU, J., HU, J., ZHOU, J., LIN, Y., WEN, J.-R. et al. (2025). Large language diffusion models. arXiv preprint. Available at arXiv:2502.09992.

ORABONA, F. (2020). Neural networks (maybe) evolved to make Adam the best optimizer. Available at https://parameterfree.com/2020/12/06/neural-network-maybe-evolved-to-make-adam-the-best-optimizer/.

OUYANG, L., WU, J., JIANG, X., ALMEIDA, D., WAINWRIGHT, C., MISHKIN, P., ZHANG, C., AGARWAL, S., SLAMA, K. et al. (2022). Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems* **35** 27730–27744.

PAN, X., YE, C., MELKONIAN, J., MA, J. W. and ZHANG, T. (2025). Daunce: Data attribution through uncertainty estimation. arXiv preprint. Available at arXiv:2505.23223.

PANG, Q., HU, S., ZHENG, W. and SMITH, V. (2024). Attacking LLM watermarks by exploiting their strengths. In *ICLR* 2024 *Workshop on Secure and Trustworthy Large Language Models*.

PARK, S. M., GEORGIEV, K., ILYAS, A., LECLERC, G. and MADRY, A. (2023). TRAK: Attributing model behavior at scale. In *International Conference on Machine Learning* 27074–27113. PMLR.

PENG, B., ALCAIDE, E., ANTHONY, Q. G., ALBALAK, A., ARCADINHO, S., BIDERMAN, S., CAO, H., CHENG, X., CHUNG, M. N. et al. (2023). RWKV: Reinventing RNNs for the Transformer era. In *The* 2023 *Conference on Empirical Methods in Natural Language Processing*.

PHAN, B., HAVASI, M., MUCKLEY, M. and ULLRICH, K. (2024). Understanding and mitigating tokenization bias in language models. arXiv preprint. Available at arXiv:2406.16829.

POLO, F. M., WEBER, L., CHOSHEN, L., SUN, Y., XU, G. and YUROCHKIN, M. (2024). tinyBenchmarks: Evaluating llms with fewer examples. In *International Conference on Machine Learning*.

RADFORD, A., NARASIMHAN, K., SALIMANS, T. and SUTSKEVER, I. (2018). Improving language understanding by generative pre-training.

RATKOVIC, M. (2025). Large language models for statistical inference: Context augmentation with applications to the two-sample problem and regression. arXiv preprint. Available at arXiv:2506.23862.

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* **22** 400–407. MR0042668 https://doi.org/10.1214/aoms/1177729586

SANTURKAR, S., DURMUS, E., LADHAK, F., LEE, C., LIANG, P. and HASHIMOTO, T. (2023). Whose opinions do language models reflect? In *International Conference on Machine Learning* 29971–30004. PMLR.

SAVANI, Y., TROCKMAN, A., FENG, Z., SCHWARZSCHILD, A., ROBEY, A., FINZI, M. and KOLTER, J. Z. (2025). Antidistillation sampling. arXiv preprint. Available at arXiv:2504.13146.

SCHEID, A., BOURSIER, E., DURMUS, A., JORDAN, M. I., MÉNARD, P., MOULINES, E. and VALKO, M. (2024). Optimal design for reward modeling in RLHF. arXiv preprint. Available at arXiv:2410.17055.

SCHMIDT, L. (2023). *Are Transformers Necessary? A Data-Centric View on Generalization. Workshop on Large Language Models and Transformers at the Simons Institute for the Theory of Computing*. Univ. California Press, Berkeley, CA.

SHI, R., ZHOU, R. and DU, S. S. (2025). The crucial role of samplers in online direct preference optimization. In *The Thirteenth International Conference on Learning Representations*.

SHUMAILOV, I., SHUMAYLOV, Z., ZHAO, Y., PAPERNOT, N., ANDERSON, R. and GAL, Y. (2024). AI models collapse when trained on recursively generated data. *Nature* **631** 755–759.

SILVER, D. and SUTTON, R. S. (2025). *Welcome to the Era of Experience*. MIT Press, Cambridge. Preprint of a chapter that will appear in the book Designing an Intelligence.

SIMONYAN, K. and ZISSERMAN, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint. Available at arXiv:1409.1556.

SRIRAMANAN, G., BHARTI, S., SADASIVAN, V. S., SAHA, S., KATTAKINDA, P. and FEIZI, S. (2024). LLM-check: Investigating detection of hallucinations in large language models. *Adv. Neural Inf. Process. Syst.* **37** 34188–34216.

SU, W. (2025). Isotropic curvature model for understanding deep learning optimization: Is gradient orthogonalization optimal? arXiv preprint. Available at arXiv:2511.00674.

SU, W. J. (2024). Envisioning future deep learning theories: Some basic concepts and characteristics. *Sci. China Inf. Sci.* **67** 203101.

SUI, Y., CHUANG, Y.-N., WANG, G., ZHANG, J., ZHANG, T., YUAN, J., LIU, H., WEN, A., ZHONG, S. et al. (2025). Stop overthinking: a survey on efficient reasoning for large language models. arXiv preprint. Available at arXiv:2503.16419.

SUTSKEVER, I. (2024). Sequence to sequence learning with neural networks: What a decade. NeurIPS 2024 Keynote. Available at https://www.youtube.com/watch?v=1yvBqasHLZs.

SUTTON, R. S. (2019). The bitter lesson. Available at http://www.incompleteideas.net/IncIdeas/BitterLesson.html. Accessed: 2024-05-24.

SWAMY, G., CHOUDHURY, S., SUN, W., WU, Z. S. and BAGNELL, J. A. (2025). All roads lead to likelihood: the value of reinforcement learning in fine-tuning. arXiv preprint. Available at arXiv:2503.01067.

THRUSH, T., POTTS, C. and HASHIMOTO, T. (2024). Improving pretraining data using perplexity correlations. arXiv preprint. Available at arXiv:2409.05816.

TIAN, X. and SHEN, X. (2025). Conditional data synthesis augmentation. arXiv preprint. Available at arXiv:2504.07426.

TU, X., ZOU, J., SU, W. and ZHANG, L. (2024). What should data science education do with large language models? *Harv. Data Sci. Rev.* **6**.

VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, Ł. and POLOSUKHIN, I. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**.

WANG, X., WEI, J., SCHUURMANS, D., LE, Q. V., CHI, E. H., NARANG, S., CHOWDHERY, A. and ZHOU, D. (2023). Self-consistency improves chain of thought reasoning in language models. In *The Eleventh International Conference on Learning Representations*.

WANG, Z., WANG, Q., ZHANG, Y., CHEN, T., ZHU, X., SHI, X. and XU, K. (2025). SConU: Selective conformal uncertainty in large language models. arXiv preprint. Available at arXiv:2504.14154.

WEI, J., WANG, X., SCHUURMANS, D., BOSMA, M., XIA, F., CHI, E., LE, Q. V. and ZHOU, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems* **35** 24824–24837.

WOLFRAM, S. and GAD-EL HAK, M. (2003). A new kind of science. *Appl. Mech. Rev.* **56** B18–B19.

XIAO, J., HOU, B., WANG, Z., JIN, R., LONG, Q., SU, W. J. and SHEN, L. (2025a). Restoring calibration for aligned large language models: A calibration-aware fine-tuning approach. In *International Conference on Machine Learning*.

XIAO, J., LI, Z., XIE, X., GETZEN, E., FANG, C., LONG, Q. and SU, W. J. (2025b). On the algorithmic bias of aligning large language models with RLHF: Preference collapse and matching regularization. *J. Amer. Statist. Assoc.* **120** 2154–2164. MR5006831 https://doi.org/10.1080/01621459.2025.2555067

XIE, C., LIN, Z., BACKURS, A., GOPI, S., YU, D., INAN, H. A., NORI, H., JIANG, H., ZHANG, H. et al. (2024). Differentially private synthetic data via foundation model APIs 2: Text. In *International Conference on Machine Learning* 54531–54560. PMLR.

XIE, S. M., PHAM, H., DONG, X., DU, N., LIU, H., LU, Y., LIANG, P. S., LE, Q. V., MA, T. et al. (2023). Doremi: Optimizing data mixtures speeds up language model pretraining. *Adv. Neural Inf. Process. Syst.* **36** 69798–69818.

YADKORI, Y. A., KUZBORSKIJ, I., GYÖRGY, A. and SZEPESVÁRI, C. (2024). To believe or not to believe your llm. arXiv preprint. Available at arXiv:2406.02543.

YANG, J., WANG, Z., LIN, Y. and ZHAO, Z. (2024). Problematic tokens: Tokenizer bias in large language models. In 2024 *IEEE International Conference on Big Data* (*BigData*) 6387–6393. IEEE.

YANG, Z., BAND, N., LI, S., CANDÈS, E. and HASHIMOTO, T. (2025). Synthetic continued pretraining. In *International Conference on Learning Representations*.

YAO, S. (2025). The second half. Available at https://ysymyth.github.io/The-Second-Half/. Online essay about AI development stages.

YAO, Y., XU, X. and LIU, Y. (2024). Large language model unlearning. *Adv. Neural Inf. Process. Syst.* **37** 105425–105475.

YE, K., ZHOU, H., ZHU, J., QUINZAN, F. and SHI, C. (2025). Robust reinforcement learning from human feedback for large language models fine-tuning. arXiv preprint. Available at arXiv:2504.03784.

YOUSEFI, M. and COLLINS, J. (2024). Learning the bitter lesson: Empirical evidence from 20 years of CVPR proceedings. In *Proceedings of the* 1*st Workshop on NLP for Science* (*NLP*4*Science*) 175–187.

YUE, Y., CHEN, Z., LU, R., ZHAO, A., WANG, Z., SONG, S. and HUANG, G. (2025). Does reinforcement learning really incentivize reasoning capacity in llms beyond the base model? arXiv preprint. Available at arXiv:2504.13837.

ZHANG, L., ROTH, A. and ZHANG, L. (2024). Fair risk control: A generalized framework for calibrating multi-group fairness risks. In *International Conference on Machine Learning* 59783–59805. PMLR.

ZHANG, R., LIN, L., BAI, Y. and MEI, S. (2024). Negative preference optimization: From catastrophic collapse to effective unlearning. In *First Conference on Language Modeling*.

ZHU, B., JORDAN, M. and JIAO, J. (2023). Principled reinforcement learning with human feedback from pairwise or k-wise comparisons. In *International Conference on Machine Learning* 43037–43067. PMLR.

ZHUANG, Y., LIU, L., SINGH, C., SHANG, J. and GAO, J. (2025). Text generation beyond discrete token sampling. arXiv preprint. Available at arXiv:2505.14827.

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 744–763
https://doi.org/10.1214/25-AOAS2080

A BAYESIAN REINFORCEMENT LEARNING FRAMEWORK FOR OPTIMIZING THE BCI-UTILITY OF P300 BRAIN-COMPUTER INTERFACES

BY BANGYAO ZHAO[1,a], YIXIN WANG[2,c], JANE E. HUGGINS[3,d] AND JIAN KANG[1,b]

[1]*Department of Biostatistics, University of Michigan,* [a]*byzhao@umich.edu,* [b]*jiankang@umich.edu*
[2]*Department of Statistics, University of Michigan,* [c]*yixinw@umich.edu*
[3]*Physical Medicine & Rehabilitation, University of Michigan,* [d]*janeh@umich.edu*

Brain-computer interfaces (BCIs) enable direct communication between the brain and computers, providing critical tools for people with disabilities to communicate with the world. The performance of BCIs is often evaluated using BCI-utility, a comprehensive metric that balances both accuracy and speed in communication. This paper introduces a Bayesian reinforcement learning framework to optimize the BCI-utility of the P300 BCI, a BCI system that identifies a user's intended character on a virtual keyboard by analyzing EEG responses to stimuli. We construct confidence scores for each character based on EEG responses and then propose a unified learning framework that explicitly maximizes BCI utility. It integrates two key components: an early stopping policy and a dynamic stimulus selection policy. The early stopping policy is optimized using an actor-critic algorithm, while a Gaussian process-based Bayesian model is developed to learn transition dynamics to guide the selection of the next stimulus. The proposed framework effectively addresses critical implementation challenges, including pauses between characters, double-target issues, and delays caused by the time required for EEG responses. Extensive simulations under varying signal-to-noise ratios (SNRs) and evaluations on recorded human EEG data demonstrate that our method significantly improves BCI-utility compared to existing approaches. This work highlights the potential of reinforcement learning to improve the performance and usability of P300 BCI systems.

REFERENCES

AKMAN AYDIN, E., BAY, Ö. F. and GÜLER, İ. (2017). A dynamic stopping algorithm for P300-based brain-computer interface systems. In *Proceedings of the International Conference on Medical and Biological Engineering* (*CMBEBIH* 2017) 723–728. Springer, Berlin.

BARTHÉLEMY, Q., CHEVALLIER, S., BERTRAND-LALO, R. and CLISSON, P. (2023). End-to-end P300 BCI using Bayesian accumulation of Riemannian probabilities. *Brain-Comput. Interfaces* **10** 50–61.

CHEN, Z. and ZHANG, X. (2017). Dynamic stopping in P300 speller with convolutional neural network. In 2017 8*th International IEEE/EMBS Conference on Neural Engineering* (*NER*) 383–386. IEEE.

FARWELL, L. A. and DONCHIN, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* **70** 510–523.

GRONDMAN, I., BUSONIU, L., LOPES, G. A. D. and BABUSKA, R. (2012). A survey of actor-critic reinforcement learning: Standard and natural policy gradients. *IEEE Trans. Syst. Man Cybern., Part C Appl. Rev.* **42** 1291–1307.

HUGGINS, J. E., MOINUDDIN, A. A., CHIODO, A. E. and WREN, P. A. (2015). What would brain-computer interface users want: Opinions and priorities of potential users with spinal cord injury. *Arch. Phys. Med. Rehabil.* **96** S38–S45.

HUGGINS, J. E., WREN, P. A. and GRUIS, K. L. (2011). What would brain-computer interface users want? Opinions and priorities of potential users with amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler.* **12** 318–324.

JAMSHIDI IDAJI, M., SHAMSOLLAHI, M. B. and HAJIPOUR SARDOUIE, S. (2017). Higher order spectral regression discriminant analysis (HOSRDA): A tensor feature reduction method for ERP detection. *Pattern Recognit.* **70** 152–162.

Key words and phrases. Model-based reinforcement learning, brain-computer interface, dynamic stimulus pattern, early stopping, Gaussian process.

JIN, J., ALLISON, B. Z., KAUFMANN, T., KÜBLER, A., ZHANG, Y., WANG, X. and CICHOCKI, A. (2012). The changing face of P300 BCIs: A comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement. *PLoS ONE* **7** e49688.

KALIKA, D., COLLINS, L. M., THROCKMORTON, C. S. and MAINSAH, B. O. (2017). Adaptive stimulus selection in ERP-based brain-computer interfaces by maximizing expected discrimination gain. In 2017 *IEEE International Conference on Systems*, *Man*, *and Cybernetics* (*SMC*) 1405–1410. IEEE.

KANG, J. and BURKARDT, J. (2022). Fast Bayesian Gaussian process regression fitting. Comprehensive R Archive Network (CRAN). Version 1.1.0.

KAPER, M., MEINICKE, P., GROSSEKATHOEFER, U., LINGNER, T. and RITTER, H. (2004). BCI competition 2003-data set IIb: Support vector machines for the P300 speller paradigm. *IEEE Trans*. *Biomed*. *Eng*. **51** 1073–1076.

KRUSIENSKI, D. J., SELLERS, E. W., CABESTAING, F., BAYOUDH, S., MCFARLAND, D. J., VAUGHAN, T. M. and WOLPAW, J. R. (2006). A comparison of classification techniques for the P300 speller. *J*. *Neural Eng*. **3** 299.

LAWHERN, V. J., SOLON, A. J., WAYTOWICH, N. R., GORDON, S. M., HUNG, C. P. and LANCE, B. J. (2018). EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. *J*. *Neural Eng*. **15** 056013.

LI, M., WU, L., LIN, F., GUO, M. and XU, G. (2022). Dual stimuli interface with logical division using local move stimuli. *Cogn*. *Neurodyn*. 1–9.

LIN, Z., SI, Y. and KANG, J. (2024). Latent subgroup identification in image-on-scalar regression. *Ann*. *Appl*. *Stat*. **18** 468–486. MR4698616 https://doi.org/10.1214/23-aoas1797

LUO, F.-M., XU, T., LAI, H., CHEN, X.-H., ZHANG, W. and YU, Y. (2024). A survey on model-based reinforcement learning. *Sci*. *China Inf*. *Sci*. **67** 121101. MR4695869 https://doi.org/10.1007/s11432-022-3696-5

MA, G., KANG, J., THOMPSON, D. E. and HUGGINS, J. E. (2023). BCI-utility metric for asynchronous P300 brain-computer interface systems. *IEEE Trans*. *Neural Syst*. *Rehabil*. *Eng*..

MA, T., HUGGINS, J. E. and KANG, J. (2021). Adaptive sequence-based stimulus selection in an ERP-based brain-computer interface by Thompson sampling in a multi-armed bandit problem. In 2021 *IEEE International Conference on Bioinformatics and Biomedicine* (*BIBM*) 3648–3655.

MOERLAND, T. M., BROEKENS, J., PLAAT, A., JONKER, C. M. et al. (2023). Model-based reinforcement learning: A survey. *Found*. *Trends Mach*. *Learn*. **16** 1–118.

PARK, J. and KIM, K.-E. (2012). A POMDP approach to optimizing P300 speller BCI paradigm. *IEEE Trans*. *Neural Syst*. *Rehabil*. *Eng*. **20** 584–594.

SAKAMOTO, Y. and AONO, M. (2009). Supervised adaptive downsampling for P300-based brain-computer interface. In 2009 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 567–570. IEEE.

SARRAF, J., PATTNAIK, P. et al. (2023). A study of classification techniques on P300 speller dataset. *Mater*. *Today Proc*. **80** 2047–2050.

SENO, B. D., MATTEUCCI, M. and MAINARDI, L. T. (2009). The utility metric: A novel method to assess the overall performance of discrete brain-computer interfaces. *IEEE Trans*. *Neural Syst*. *Rehabil*. *Eng*. **18** 20–28.

SUTTON, R. S., MCALLESTER, D., SINGH, S. and MANSOUR, Y. (1999). Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems* (S. Solla, T. Leen and K. Müller, eds.) **12**. MIT Press, Cambridge.

THOMPSON, D. E., GRUIS, K. L. and HUGGINS, J. E. (2014). A plug-and-play brain-computer interface to operate commercial assistive technology. *Disabil*. *Rehabil*., *Assist*. *Technol*. **9** 144–150.

THROCKMORTON, C. S., COLWELL, K. A., RYAN, D. B., SELLERS, E. W. and COLLINS, L. M. (2013). Bayesian approach to dynamically controlling data collection in P300 spellers. *IEEE Trans*. *Neural Syst*. *Rehabil*. *Eng*. **21** 508–517.

TOWNSEND, G., LAPALLO, B. K., BOULAY, C. B., KRUSIENSKI, D. J., FRYE, G. E., HAUSER, C. K., SCHWARTZ, N. E., VAUGHAN, T. M., WOLPAW, J. R. et al. (2010). A novel P300-based brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns. *Clin*. *Neurophysiol*. **121** 1109–1120.

WU, B., GUO, Y. and KANG, J. (2024). Bayesian spatial blind source separation via the thresholded Gaussian process. *J*. *Amer*. *Statist*. *Assoc*. **119** 422–433. MR4713903 https://doi.org/10.1080/01621459.2022.2123336

ZHAO, B., WANG, Y., HUGGINS, J. E and KANG, J. (2026). Supplement to “A Bayesian Reinforcement Learning Framework for Optimizing the BCI-utility of P300 Brain-Computer Interfaces.” https://doi.org/10.1214/25-AOAS2080SUPPA, https://doi.org/10.1214/25-AOAS2080SUPPB

ZHOU, X., HAO, B., LATTIMORE, T., KANG, J. and LI, L. (2024). Sequential best-arm identification with application to P300 speller. *Trans*. *Mach*. *Learn*. *Res*..

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 764–788
https://doi.org/10.1214/25-AOAS2113
© Institute of Mathematical Statistics, 2026

STATISTICAL INFERENCE FOR COVARIATE-ADJUSTED AND INTERPRETABLE GENERALIZED LATENT FACTOR MODEL WITH APPLICATION TO TESTING FAIRNESS

BY JING OUYANG[1,a], CHENGYU CUI[2,b], KEAN MING TAN[2,c] AND GONGJUN XU[2,d]

[1]*Faculty of Business and Economics, University of Hong Kong,* [a]*jingoy@hku.hk*
[2]*Department of Statistics, University of Michigan,* [b]*chyc@umich.edu,* [c]*keanming@umich.edu,* [d]*gongjun@umich.edu*

Latent variable models are popularly used to measure latent embedding factors from large-scale assessment data. Beyond understanding these latent factors, the covariate effect on responses controlling for latent factors is also of great scientific interest and has wide applications, such as evaluating the fairness of educational testing, where the covariate effect reflects whether a test question is biased toward certain individual characteristics (e.g., gender and race), taking into account their latent abilities. However, the large sample sizes and high-dimensional responses pose challenges to developing efficient methods and drawing valid inferences. Moreover, to accommodate the commonly encountered discrete responses, generalized latent factor models are often assumed, adding further complexity. To address these challenges, we consider a covariate-adjusted generalized factor model and develop novel and interpretable conditions to address the identifiability issue. Based on the identifiability conditions, we propose a joint maximum likelihood estimation method and establish estimation consistency and asymptotic normality results for the covariate effects. Furthermore, we derive estimation and inference results for latent factors and the factor loadings. We illustrate the finite sample performance of the proposed method through extensive numerical studies and an educational assessment dataset from the Programme for International Student Assessment (PISA).

REFERENCES

AKAIKE, H. (1987). Factor analysis and AIC. *Psychometrika* **52** 317–332. MR0914459 https://doi.org/10.1007/BF02294359

ALLMAN, E. S., MATIAS, C. and RHODES, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.* **37** 3099–3132. MR2549554 https://doi.org/10.1214/09-AOS689

ASPAROUHOV, T. and MUTHÉN, B. (2014). Multiple-group factor analysis alignment. *Struct. Equ. Model.* **21** 495–508. MR3268570 https://doi.org/10.1080/10705511.2014.919210

BAI, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71** 135–171. MR1956857 https://doi.org/10.1111/1468-0262.00392

BAI, J. and LI, K. (2012). Statistical analysis of factor models of high dimension. *Ann. Statist.* **40** 436–465. MR3014313 https://doi.org/10.1214/11-AOS966

BALART, P. and OOSTERVEEN, M. (2019). Females show more sustained performance during test-taking than males. *Nat. Commun.* **10** 3798.

BAUER, D. J., BELZAK, W. C. M. and COLE, V. T. (2020). Simplifying the assessment of measurement invariance over multiple background variables: Using regularized moderated nonlinear factor analysis to detect differential item functioning. *Struct. Equ. Model.* **27** 43–55. MR4054417 https://doi.org/10.1080/10705511.2019.1642754

BECHGER, T. M. and MARIS, G. (2015). A statistical test for differential item pair functioning. *Psychometrika* **80** 317–340. MR3353960 https://doi.org/10.1007/s11336-014-9408-y

BELZAK, W. and BAUER, D. J. (2020). Improving the assessment of measurement invariance: Using regularization to select anchor items and identify differential item functioning. *Psychol. Methods* **25** 673–690.

BING, X., CHENG, W., FENG, H. and NING, Y. (2024). Inference in high-dimensional multivariate response regression with hidden variables. *J. Amer. Statist. Assoc.* **119** 2066–2077. MR4797923 https://doi.org/10.1080/01621459.2023.2241701

Key words and phrases. Generalized latent factor model, identifiability, large-scale assessment.

BIRNBAUM, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In *Statistical Theories of Mental Test Scores* (F. M. Lord and M. R. Novick, eds.) 395–479. Addison-Wesley, Reading, MA.

BOLLMANN, S., BERGER, M. and TUTZ, G. (2018). Item-focused trees for the detection of differential item functioning in partial credit models. *Educ. Psychol. Meas.* **78** 781–804.

BROWN, T. A., MOORE, M. T. et al. (2012). Confirmatory factor analysis. In *Handbook of Structural Equation Modeling* **361** 379.

CAI, T. and ZHOU, W.-X. (2013). A max-norm constrained minimization approach to 1-bit matrix completion. *J. Mach. Learn. Res.* **14** 3619–3647. MR3159403

CANDELL, G. L. and DRASGOW, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Appl. Psychol. Meas.* **12** 253–260.

CAO, M., TAY, L. and LIU, Y. (2017). A Monte Carlo study of an iterative Wald test procedure for DIF analysis. *Educ. Psychol. Meas.* **77** 104–118.

CARVALHO, C. M., CHANG, J., LUCAS, J. E., NEVINS, J. R., WANG, Q. and WEST, M. (2008). High-dimensional sparse factor modeling: Applications in gene expression genomics. *J. Amer. Statist. Assoc.* **103** 1438–1456. MR2655722 https://doi.org/10.1198/016214508000000869

CATTELL, R. B. (1966). The scree test for the number of factors. *Multivar. Behav. Res.* **1** 245–276.

CHEN, M., FERNÁNDEZ-VAL, I. and WEIDNER, M. (2021). Nonlinear factor models for network and panel data. *J. Econometrics* **220** 296–324. MR4201492 https://doi.org/10.1016/j.jeconom.2020.04.004

CHEN, Y., LI, C., OUYANG, J. and XU, G. (2023a). Statistical inference for noisy incomplete binary matrix. *J. Mach. Learn. Res.* **24** 1–66. MR4582517

CHEN, Y., LI, C., OUYANG, J. and XU, G. (2023b). DIF statistical inference without knowing anchoring items. *Psychometrika* **88** 1097–1122. MR4668563 https://doi.org/10.1007/s11336-023-09930-9

CHEN, Y. and LI, X. (2022). Determining the number of factors in high-dimensional generalized latent factor models. *Biometrika* **109** 769–782. MR4472847 https://doi.org/10.1093/biomet/asab044

CHEN, Y., LI, X., LIU, J. and YING, Z. (2025). Item response theory—a statistical framework for educational and psychological measurement. *Statist. Sci.* **40** 167–194. MR4915103 https://doi.org/10.1214/23-STS896

CHEN, Y., LI, X. and ZHANG, S. (2019). Joint maximum likelihood estimation for high-dimensional exploratory item factor analysis. *Psychometrika* **84** 124–146. MR3910911 https://doi.org/10.1007/s11336-018-9646-5

CHEN, Y., LI, X. and ZHANG, S. (2020). Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *J. Amer. Statist. Assoc.* **115** 1756–1770. MR4189755 https://doi.org/10.1080/01621459.2019.1635485

COLLINS, L. M. and L., S. T. (2009). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. Wiley, New York.

COLLINS, M., DASGUPTA, S. and SCHAPIRE, R. E. (2002). A generalization of principal components analysis to the exponential family. *Adv. Neural Inf. Process. Syst.* **14** 617–624.

CUI, C. and XU, G. (2025). Identifiability and inference for generalized latent factor models. arXiv preprint. Available at arXiv:2508.05866.

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

DORANS, N. J. and KULICK, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *J. Educ. Meas.* **23** 355–368.

DU, J.-H., WASSERMAN, L. and ROEDER, K. (2025). Simultaneous inference for generalized linear models with unmeasured confounders. *J. Amer. Statist. Assoc.* **120** 1945–1959. MR4973909 https://doi.org/10.1080/01621459.2025.2485379

FAN, J., LIAO, Y. and MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 603–680. MR3091653 https://doi.org/10.1111/rssb.12016

FIDALGO, A., MELLENBERGH, G. J. and MUÑIZ, J. (2000). Effects of amount of DIF, test length, and purification type on robustness and power of Mantel-Haenszel procedures. *MPR Online* **5** 43–53.

FRICHOT, E., SCHOVILLE, S. D., BOUCHARD, G. and FRANÇOIS, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol. Biol. Evol.* **30** 1687–1699.

FRICK, H., STROBL, C. and ZEILEIS, A. (2015). Rasch mixture models for DIF detection: A comparison of old and new score specifications. *Educ. Psychol. Meas.* **75** 208–234.

GERARD, D. and STEPHENS, M. (2020). Empirical Bayes shrinkage and false discovery rate estimation, allowing for unwanted variation. *Biostatistics* **21** 15–32. MR4043843 https://doi.org/10.1093/biostatistics/kxy029

GOLDBERGER, A. S. (1972). Structural equation methods in the social sciences. *Econometrica* **40** 979–1001. MR0327267 https://doi.org/10.2307/1913851

GU, Y. and XU, G. (2020). Partial identifiability of restricted latent class models. *Ann. Statist.* **48** 2082–2107. MR4134787 https://doi.org/10.1214/19-AOS1878

HABERMAN, S. J. (1977). Maximum likelihood estimates in exponential response models. *Ann. Statist.* **5** 815–841. MR0501540

HAMBLETON, R. K. and SWAMINATHAN, H. (2013). *Item Response Theory: Principles and Applications*. Springer, Berlin.

HE, X. and SHAO, Q.-M. (2000). On parameters of increasing dimensions. *J. Multivariate Anal.* **73** 120–135. MR1766124 https://doi.org/10.1006/jmva.1999.1873

HOLLAND, P. W. and WAINER, H. (2012). *Differential Item Functioning*. Routledge, London.

HORN, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* **30** 179–185.

JÖRESKOG, K. G. and GOLDBERGER, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *J. Amer. Statist. Assoc.* **70** 631–639. MR0395057

KAISER, H. F. (1960). The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* **20** 141–151.

KOPF, J., ZEILEIS, A. and STROBL, C. (2015). A framework for anchor methods and an iterative forward approach for DIF detection. *Appl. Psychol. Meas.* **39** 83–103.

KORHONEN, P., NORDHAUSEN, K. and TASKINEN, S. (2024). A review of generalized linear latent variable models and related computational approaches. *Wiley Interdiscip. Rev.: Comput. Stat.* **16** Paper No. e70005, 24. MR4829001 https://doi.org/10.1002/wics.70005

LAM, C. and YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** 694–726. MR2933663 https://doi.org/10.1214/12-AOS970

LAM, C., YAO, Q. and BATHIA, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika* **98** 901–918. MR2860332 https://doi.org/10.1093/biomet/asr048

LEEK, J. T. and STOREY, J. D. (2008). A general framework for multiple testing dependence. *Proc. Natl. Acad. Sci. USA* **105** 18718–18723.

MANTEL, N. and HAENSZEL, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22** 719–748.

MAY, H. (2006). A multilevel Bayesian item response theory method for scaling socioeconomic status in international studies of education. *J. Educ. Behav. Stat.* **31** 63–79.

MELLENBERGH, G. J. (1994). Generalized linear item response theory. *Psychol. Bull.* **115** 300–307.

MILLSAP, R. E. (2012). *Statistical Approaches to Measurement Invariance*. Routledge, New York, NY.

MOUSTAKI, I. (2003). A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *Br. J. Math. Stat. Psychol.* **56** 337–357. MR2101773 https://doi.org/10.1348/000711003770480075

MOUSTAKI, I., JÖRESKOG, K. G. and MAVRIDIS, D. (2004). Factor models for ordinal variables with covariate effects on the manifest and latent variables: A comparison of LISREL and IRT approaches. *Struct. Equ. Model.* **11** 487–513. MR2062113 https://doi.org/10.1207/s15328007sem1104_1

MOUSTAKI, I. and KNOTT, M. (2000). Generalized latent trait models. *Psychometrika* **65** 391–411. MR1792703 https://doi.org/10.1007/BF02296153

MUTHEN, B. (1985). A method for studying the homogeneity of test items with respect to other relevant variables. *J. Educ. Stat.* **10** 121–132.

MUTHÉN, B., ASPAROUHOV, T. et al. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Notes* **4** 1–22.

MUTHEN, B., KAO, C. F. and BURSTEIN, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement test items. *J. Educ. Meas.* **28** 1–22.

MUTHÉN, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika* **54** 557–585. MR1041525 https://doi.org/10.1007/BF02296397

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

NIKU, J., HUI, F. K., TASKINEN, S. and WARTON, D. I. (2019). gllvm: Fast analysis of multivariate abundance data with generalized linear latent variable models in R. *Methods Ecol. Evol.* **10** 2173–2182.

OECD (2019). *PISA* 2018 *Assessment and Analytical Framework*. *PISA*. OECD Publishing, Paris.

OECD (2020). *PISA* 2018 *Technical Report*. *PISA*. OECD Publishing, Paris.

OORT, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Struct. Equ. Model.* **5** 107–124.

OUYANG, J., CUI, C., TAN, K. M and XU, G. (2026). Supplement to “Statistical inference for covariate-adjusted and interpretable generalized latent factor model with application to testing fairness.” https://doi.org/10.1214/25-AOAS2113SUPPA, https://doi.org/10.1214/25-AOAS2113SUPPB

OUYANG, J., TAN, K. M. and XU, G. (2023). High-dimensional inference for generalized linear models with hidden confounding. *J. Mach. Learn. Res.* **24** 1–61. MR4664733

PARK, Y. S., XING, K. and LEE, Y.-S. (2018). Explanatory cognitive diagnostic models: Incorporating latent and observed predictors. *Appl. Psychol. Meas.* **42** 376–392.

QUINN, D. M. and COOC, N. (2015). Science achievement gaps by gender and race/ethnicity in elementary and middle school: Trends and predictors. *Educ. Res.* **44** 336–346.

RABE-HESKETH, S., SKRONDAL, A. and PICKLES, A. (2004). GLLAMM manual.

REBOUSSIN, B. A., IP, E. H. and WOLFSON, M. (2008). Locally dependent latent class models with covariates: An application to under-age drinking in the USA. *J. R. Stat. Soc., Ser. A* **171** 877–897. MR2530291 https://doi.org/10.1111/j.1467-985X.2008.00544.x

RECKASE, M. D. (2009). *Multidimensional Item Response Theory*. Springer, New York, NY.

RIJMEN, F., TUERLINCKX, F., DE BOECK, P. and KUPPENS, P. (2003). A nonlinear mixed model framework for item response theory. *Psychol. Methods* **8** 185–205.

SCHLEICHER, A. (2019). *PISA* 2018: *Insights and Interpretations*. OECD Publishing.

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

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

SOARES, T. M., GONÇALVES, F. B. and GAMERMAN, D. (2009). An integrated Bayesian model for DIF analysis. *J. Educ. Behav. Stat.* **34** 348–377.

STEENKAMP, J.-B. E. and BAUMGARTNER, H. (1998). Assessing measurement invariance in cross-national consumer research. *J. Consum. Res.* **25** 78–90.

STOCK, J. H. and WATSON, M. W. (2002). Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.* **97** 1167–1179. MR1951271 https://doi.org/10.1198/016214502388618960

STOCK, J. H. and WATSON, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions. In *Macroeconomics* (J. B. Taylor and H. Uhlig, eds.). *Handbook of Macroeconomics* **2** 415–525. Elsevier, Amsterdam.

STROBL, C., KOPF, J., KOHLER, L., OERTZEN, T. V. and ZEILEIS, A. (2021). Anchor point selection: Scale alignment based on an inequality criterion. *Appl. Psychol. Meas.* **45** 214–230.

TAY, L., HUANG, Q. and VERMUNT, J. K. (2016). Item response theory with covariates (IRT-C) assessing item recovery and differential item functioning for the three-parameter logistic model. *Educ. Psychol. Meas.* **76** 22–42.

THISSEN, D. (1988). Use of item response theory in the study of group differences in trace lines. In *Test Validity* (H. E. Wainer and H. I. Braun, eds.) 147–172. Erlbaum, Mahwah, NJ.

TOCH, T. (1984). Test Organization, Insurance Firm Settle Bias Suit. Education Week. https://www.edweek.org/education/test-organization-insurance-firm-settle-bias-suit/1984/12.

TUTZ, G. and BERGER, M. (2016). Item-focussed trees for the identification of items in differential item functioning. *Psychometrika* **81** 727–750. MR3535055 https://doi.org/10.1007/s11336-015-9488-3

VERMUNT, J. K. and MAGIDSON, J. (2005). *Latent GOLD® Choice* 4.0 *User's Manual*. Statistical Innovations Inc., Belmont, MA.

VON DAVIER, M. (2008). A general diagnostic model applied to language testing data. *Br. J. Math. Stat. Psychol.* **61** 287–307. MR2649038 https://doi.org/10.1348/000711007X193957

WALLIN, G., CHEN, Y. and MOUSTAKI, I. (2024). DIF analysis with unknown groups and anchor items. *Psychometrika* **89** 267–295. MR4740705 https://doi.org/10.1007/s11336-024-09948-7

WANG, C., ZHU, R. and XU, G. (2023). Using lasso and adaptive lasso to identify DIF in multidimensional 2PL models. *Multivar. Behav. Res.* **58** 387–407.

WANG, F. (2022). Maximum likelihood estimation and inference for high dimensional generalized factor models with application to factor-augmented regressions. *J. Econometrics* **229** 180–200. MR4414018 https://doi.org/10.1016/j.jeconom.2020.11.002

WANG, J., ZHAO, Q., HASTIE, T. and OWEN, A. B. (2017). Confounder adjustment in multiple hypothesis testing. *Ann. Statist.* **45** 1863–1894. MR3718155 https://doi.org/10.1214/16-AOS1511

XU, G. (2017). Identifiability of restricted latent class models with binary responses. *Ann. Statist.* **45** 675–707. MR3650397 https://doi.org/10.1214/16-AOS1464

YUAN, K.-H., LIU, H. and HAN, Y. (2021). Differential item functioning analysis without a priori information on anchor items: QQ plots and graphical test. *Psychometrika* **86** 345–377. MR4291692 https://doi.org/10.1007/s11336-021-09746-5

ZELLNER, A. (1970). Estimation of regression relationships containing unobservable independent variables. *Internat. Econom. Rev.* **11** 441–454.

ZWICK, R., THAYER, D. T. and LEWIS, C. (2000). Using loss functions for DIF detection: An empirical Bayes approach. *J. Educ. Behav. Stat.* **25** 225–247.

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 789–808
https://doi.org/10.1214/25-AOAS2123

RANKING AND SELECTION IN LARGE-SCALE INFERENCE OF HETEROSCEDASTIC UNITS

BY BOWEN GANG[1,a], LUELLA FU[2,b], GARETH M. JAMES[3,c] AND WENGUANG SUN[4,d]

[1]*Department of Statistics and Data Science, Fudan University,* [a]*bgang@fudan.edu.cn*
[2]*Department of Mathematics, San Francisco State University,* [b]*luella@sfsu.edu*
[3]*Goizueta Business School, Emory University,* [c]*gareth@emory.edu*
[4]*Center for Data Science, Zhejiang University,* [d]*wgsun@zju.edu.cn*

The allocation of limited resources to a large number of potential candidates presents a pervasive challenge. In the context of ranking and selecting top candidates from heteroscedastic units, conventional methods often result in overrepresentations of subpopulations, and this issue is further exacerbated in large-scale settings where thousands of candidates are considered simultaneously. In particular, we consider this problem in ranking schools based on socioeconomic performance gaps in standardized testing. To address this challenge, we propose a new multiple comparison framework that incorporates a modified power notion to prioritize the selection of important effects and employs a novel ranking metric to assess the relative importance of units. We develop both oracle and data-driven algorithms and demonstrate their effectiveness in controlling the error rates and achieving optimality. We evaluate the numerical performance of our proposed method using simulated and real data. The results show that our framework enables a more balanced selection of effects that are both statistically significant and practically important and results in an objective and relevant ranking scheme that is well-suited to practical scenarios.

REFERENCES

BANERJEE, T., FU, L. J., JAMES, G. M. and SUN, W. (2020). Nonparametric empirical Bayes estimation on heterogeneous data. arXiv preprint. Available at arXiv:2002.12586.

BASU, P., CAI, T. T., DAS, K. and SUN, W. (2018). Weighted false discovery rate control in large-scale multiple testing. *J. Amer. Statist. Assoc.* **113** 1172–1183. PMID: 31011234. https://doi.org/10.1080/01621459.2017.1336443

BECHHOFER, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25** 16–39. https://doi.org/10.1214/aoms/1177728845

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, Methodol.* **57** 289–300.

BOYD, S., CORTES, C., MOHRI, M. and RADOVANOVIC, A. (2012). Accuracy at the top. In *Advances in Neural Information Processing Systems* (F. Pereira, C. J. Burges, L. Bottou and K. Q. Weinberger, eds.) **25**. Curran Associates, Red Hook.

CAI, T. T. and SUN, W. (2009). Simultaneous testing of grouped hypotheses: Finding needles in multiple haystacks. *J. Amer. Statist. Assoc.* **104** 1467–1481.

CAI, T. T., SUN, W. and WANG, W. (2019). Covariate-assisted ranking and screening for large-scale two-sample inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 187–234.

CHEN, C.-H., LIN, J., YÜCESAN, E. and CHICK, S. E. (2000). Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Discrete Event Dyn. Syst.* **10** 251–270. https://doi.org/10.1023/A:1008349927281

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

EFRON, B. (2011). Tweedie's formula and selection bias. *J. Amer. Statist. Assoc.* **106** 1602–1614.

Key words and phrases. Compound decision theory, composite null hypotheses, deconvolution estimates, empirical Bayes, false discovery rate, weighted multiple testing.

EFRON, B. (2012). *Large-Scale Inference*: *Empirical Bayes Methods for Estimation*, *Testing*, *and Prediction* 1. Cambridge Univ. Press, Cambridge.

EFRON, B. (2016). Empirical Bayes deconvolution estimates. *Biometrika* **103** 1–20.

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.

FOSTER, D. P. and STINE, R. A. (2008). α-Investing: A procedure for sequential control of expected false discoveries. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 429–444.

FU, L., GANG, B., JAMES, G. M. and SUN, W. (2022). Heteroscedasticity-adjusted ranking and thresholding for large-scale multiple testing. *J. Amer. Statist. Assoc.* **117** 1028–1040.

GANG, B. and BANERJEE, T. (2025). Large-scale multiple testing of composite null hypotheses under heteroskedasticity. *Biometrika*. asaf007.

GANG, B., FU, L., JAMES, G. M and SUN, W. (2026). Supplement to "Ranking and Selection in Large-Scale Inference of Heteroscedastic Units." https://doi.org/10.1214/25-AOAS2123SUPPA, https://doi.org/10.1214/25-AOAS2123SUPPB

GANG, B., SUN, W. and WANG, W. (2023). Structure–adaptive sequential testing for online false discovery rate control. *J. Amer. Statist. Assoc.* **118** 732–745.

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. https://doi.org/10.1214/009053604000000283

GOEL, P. K. and RUBIN, H. (1977). On selecting a subset containing the best population-a Bayesian approach. *Ann. Statist.* **5** 969–983.

GOLDWASSER, J., FITHIAN, W. and HOOKER, G. (2025). Gaussian Rank Verification. Available at arXiv:2501.14142. https://doi.org/10.48550/arXiv.2501.14142

GU, J. and KOENKER, R. (2017a). Unobserved heterogeneity in income dynamics: An empirical Bayes perspective. *J. Bus. Econom. Statist.* **35** 1–16.

GU, J. and KOENKER, R. (2017b). Empirical bayesball remixed: Empirical Bayes methods for longitudinal data. *J. Appl. Econometrics* **32** 575–599.

GU, J. and KOENKER, R. (2023). Invidious comparisons: Ranking and selection as compound decisions. *Econometrica* **91** 1–41.

GUPTA, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics* **7** 225–245.

HANUSHEK, E. A. (2011). The economic value of higher teacher quality. *Econ. Educ. Rev.* **30** 466–479. https://doi.org/10.1016/j.econedurev.2010.12.006

HE, L., SARKAR, S. K. and ZHAO, Z. (2015). Capturing the severity of type II errors in high-dimensional multiple testing. *J. Multivariate Anal.* **142** 106–116.

HENDERSON, N. C. and NEWTON, M. A. (2016). Making the cut: Improved ranking and selection for large-scale inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 781.

HSU, J. C. (1992). Stepwise multiple comparisons with the best. *J. Statist. Plann. Inference* **33** 197–204. https://doi.org/10.1016/0378-3758(92)90067-3

HUNG, K. and FITHIAN, W. (2019). Rank verification for exponential families. *Ann. Statist.* **47** 758–782.

JIANG, W. and ZHANG, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.* **37** 1647–1684.

KAMIŃSKI, B. and SZUFEL, P. (2018). On parallel policies for ranking and selection problems. *J. Appl. Stat.* **45** 1690–1713. https://doi.org/10.1080/02664763.2017.1390555

KOENKER, R. and MIZERA, I. (2014). Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *J. Amer. Statist. Assoc.* **109** 674–685.

KWON, Y. and ZHAO, Z. (2023). On F-modelling-based empirical Bayes estimation of variances. *Biometrika* **110** 69–81.

LIN, R., LOUIS, T. A., PADDOCK, S. M. and RIDGEWAY, G. (2006). Loss function based ranking in two-stage, hierarchical models. *Bayesian Anal.* **1** 915–946. https://doi.org/10.1214/06-BA130

LIN, R., LOUIS, T. A., PADDOCK, S. M. and RIDGEWAY, G. (2009). Ranking USRDS provider specific SMRs from 1998–2001. *Health Serv. Outcomes Res. Methodol.* **1** 22–38. https://doi.org/10.1007/s10742-008-0040-0

LUO, J., HONG, L. J., NELSON, B. L. and WU, Y. (2015). Fully sequential procedures for large-scale ranking-and-selection problems in parallel computing environments. *Oper. Res.* **63** 1177–1194.

MOSTELLER, F. (1948). A k-sample slippage test for an extreme population. *Ann. Math. Statist.* **19** 58–65. https://doi.org/10.1214/aoms/1177730290

NI, E. C., CIOCAN, D. F., HENDERSON, S. G. and HUNTER, S. R. (2017). Efficient ranking and selection in parallel computing environments. *Oper. Res.* **65** 821–836. https://doi.org/10.1287/opre.2016.1577

PANCHAPAKESAN, S. (1971). On a subset selection procedure for the most probable event in a multinomial distribution. In *Statistical Decision Theory and Related Topics* (S. S. Gupta and J. Yackel, eds.) 275–298. Academic Press, San Diego. https://doi.org/10.1016/B978-0-12-307550-5.50019-8

PAULSON, E. (1949). A multiple decision procedure for certain problems in the analysis of variance. *Ann. Math. Statist.* **20** 95–98. https://doi.org/10.1214/aoms/1177730094

ROGOSA, D. (2003). Accuracy of API index and school base report elements: 2003 Academic Performance Index. California Department of Education. Report.

SHANI, G. and GUNAWARDANA, A. (2011). *Evaluating Recommendation Systems in Recommender Systems Handbook* 257–297. Springer US, Boston, MA. https://doi.org/10.1007/978-0-387-85820-3_8

STOREY, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 479–498.

SUN, W. and CAI, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.* **102** 901–912.

SUN, W. and MCLAIN, A. C. (2012). Multiple testing of composite null hypotheses in heteroscedastic models. *J. Amer. Statist. Assoc.* **107** 673–687.

TANG, J., HU, X. and LIU, H. (2013). Social recommendation: A review. *Soc. Netw. Anal. Min.* **3** 1113–1133.

WAND, M. P. and JONES, M. C. (1994). *Kernel Smoothing. Chapman and Hall CRC Monographs on Statistics and Applied Probability* **60**. CRC Press, Boca Raton.

WEINSTEIN, A., MA, Z., BROWN, L. D. and ZHANG, C.-H. (2018). Group-linear empirical Bayes estimates for a heteroscedastic normal mean. *J. Amer. Statist. Assoc.* 1–13. https://doi.org/10.1080/01621459.2017.1280406

XIE, X., KOU, S. and BROWN, L. D. (2012). SURE estimates for a heteroscedastic hierarchical model. *J. Amer. Statist. Assoc.* **107** 1465–1479.

ZHONG, Y., LIU, S., LUO, J. and HONG, L. J. (2022). Speeding up paulson's procedure for large-scale problems using parallel computing. *INFORMS J. Comput.* **34** 586–606. https://doi.org/10.1287/ijoc.2020.1054

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 809–832
https://doi.org/10.1214/25-AOAS2126

RANDOM FORESTS AND MIXED EFFECTS RANDOM FORESTS FOR SMALL AREA ESTIMATION OF GENERAL PARAMETERS: A POVERTY MAPPING CASE STUDY IN MOZAMBIQUE

BY PATRICK KRENNMAIR[1,a], NORA WÜRZ[2,b], TIMO SCHMID[2,c] AND NIKOS TZAVIDIS[3,d]

[1]*Institute of Statistics and Econometrics, Freie Universität Berlin,* [a]*p.krennmair@gmail.com*

[2]*Institute of Statistics, Otto-Friedrich-Universität Bamberg,* [b]*nora.wuerz@uni-bamberg.de,* [c]*timo.schmid@uni-bamberg.de*

[3]*Department of Social Statistics and Demography & Southampton Statistical Sciences Research Institute, University of Southampton,* [d]*n.tzavidis@soton.ac.uk*

Use of standard random forests may not guarantee reliable small area estimates unless a rich source of predictors explains the between-area heterogeneity. We propose mixed effects random forests with area random effects for small area estimation of general parameters. A new fitting algorithm with an embedded bootstrap-bias correction for the random forest residual variance is presented. Point estimators of small area parameters are constructed using a smearing estimator of the area-specific distribution function. Nonparametric block bootstrap is used for MSE estimation. The methodology is evaluated using household consumption data from Mozambique to derive district estimates of head count ratio and poverty gap. Comparisons to the empirical best predictor under a linear mixed model and to a synthetic estimator under the random forest are presented. Estimates are further contrasted to 2023 World Bank estimates and to design-unbiased direct estimates. The results show: (a) the advantages from including random effects in random forests, (b) the importance of data transformations for machine learning methods, (c) robustness properties of random forest-type methods, and (d) the importance of bias correcting the naive estimator of the random forest residual variance. Our conclusions demonstrate that a black-box approach to using machine learning methods should be avoided.

REFERENCES

ANDERSON, W., GUIKEMA, S., ZAITCHIK, B. and PAN, W. (2014). Methods for estimating population density in data-limited areas: Evaluating regression and tree-based models in Peru. *PLoS ONE* **9**.

BATTESE, G. E., HARTER, R. M. and FULLER, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83** 28–36.

BILTON, P., JONES, G., GANESH, S. and HASLETT, S. (2017). Classification trees for poverty mapping. *Comput. Statist. Data Anal.* **115** 53–66. MR3683128 https://doi.org/10.1016/j.csda.2017.05.009

BILTON, P., JONES, G., GANESH, S. and HASLETT, S. (2020). Regression trees for poverty mapping. *Aust. N. Z. J. Stat.* **62** 426–443. MR4221961 https://doi.org/10.1111/anzs.12312

BOSA, K., BEAUMONT, J. F., BOCCI, C. and SOMBO, S. (2024). The use of random forests in small area estimation. In *Proceedings of the CANSSI Workshop, Ottawa, July* 2024.

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

CHAMBERS, R. and CHANDRA, H. (2013). A random effect block bootstrap for clustered data. *J. Comput. Graph. Statist.* **22** 452–470. MR3173724 https://doi.org/10.1080/10618600.2012.681216

CHAMBERS, R., CHANDRA, H., SALVATI, N. and TZAVIDIS, N. (2014). Outlier robust small area estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 47–69. MR3153933 https://doi.org/10.1111/rssb.12019

CHAMBERS, R. and TZAVIDIS, N. (2006). *M*-quantile models for small area estimation. *Biometrika* **93** 255–268. MR2278081 https://doi.org/10.1093/biomet/93.2.255

CHAMBERS, R. L., DORFMAN, A. H. and WEHRLY, T. E. (1993). Bias robust estimation in finite populations using nonparametric calibration. *J. Amer. Statist. Assoc.* **88** 268–277. MR1212490

Key words and phrases. Bootstrap bias correction, machine learning, mean squared error, mixed models, official statistics, statistical learning.

CHAMBERS, R. L. and DUNSTAN, R. (1986). Estimating distribution functions from survey data. *Biometrika* **73** 597–604. MR0897851 https://doi.org/10.1093/biomet/73.3.597

CHI, G., FANG, H., CHATTERJEE, S. and BLUMENSTOCK, E. J. (2022). Microestimates of wealth for all low and middle income countries. *Proc. Natl. Acad. Sci. USA* **119** 381–399.

DAGDOUG, M., GOGA, C. and HAZIZA, D. (2023). Model-assisted estimation through random forests in finite population sampling. *J. Amer. Statist. Assoc.* **118** 1234–1251. MR4595490 https://doi.org/10.1080/01621459.2021.1987250

DATTA, G. S. and LAHIRI, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statist. Sinica* **10** 613–627. MR1769758

DIALLO, M. S. and RAO, J. N. K. (2018). Small area estimation of complex parameters under unit-level models with skew-normal errors. *Scand. J. Stat.* **45** 1092–1116. MR3884901 https://doi.org/10.1111/sjos.12336

DUAN, N. (1983). Smearing estimate: A nonparametric retransformation method. *J. Amer. Statist. Assoc.* **78** 605–610. MR0721207

EFRON, B. (2014). Estimation and accuracy after model selection. *J. Amer. Statist. Assoc.* **109** 991–1007. MR3265671 https://doi.org/10.1080/01621459.2013.823775

EFRON, B. (2020). Prediction, estimation, and attribution. *J. Amer. Statist. Assoc.* **115** 636–655. MR4107663 https://doi.org/10.1080/01621459.2020.1762613

EUROSTAT (2019). DataCollection: precision level DCF. Eurostat, Luxembourg. Available from https://datacollection.jrc.ec.europa.eu/wordef/precision-level-dcf. [Accessed: 06.2019].

FOSTER, J., GREER, J. and THORBECKE, E. (1984). A class of decomposable poverty measures. *Econometrica* **52** 761–766.

GONZÁLEZ-MANTEIGA, W., LOMBARDÍA, M. J., MOLINA, I., MORALES, D. and SANTAMARÍA, L. (2008). Analytic and bootstrap approximations of prediction errors under a multivariate Fay–Herriot model. *Comput. Statist. Data Anal.* **52** 5242–5252. MR2526589 https://doi.org/10.1016/j.csda.2008.04.031

GRAF, M., MARÍN, J. M. and MOLINA, I. (2019). A generalized mixed model for skewed distributions applied to small area estimation. *TEST* **28** 565–597. MR3962067 https://doi.org/10.1007/s11749-018-0594-2

GREENWELL, B. M. (2017). pdp: An R package for constructing partial dependence plots. *R J.* **9** 421–436.

GREENWELL, B. M., BOEHMKE, B. and GRAY, B. (2020). Variable importance plots: An introduction to the vip package. *R J.* **12** 343–366.

HAJJEM, A., BELLAVANCE, F. and LAROCQUE, D. (2014). Mixed-effects random forest for clustered data. *J. Stat. Comput. Simul.* **84** 1313–1328. MR3169395 https://doi.org/10.1080/00949655.2012.741599

HALL, P. and MAITI, T. (2006). On parametric bootstrap methods for small area prediction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 221–238. MR2188983 https://doi.org/10.1111/j.1467-9868.2006.00541.x

HARMENING, S., KREUTZMANN, A.-K., SCHMIDT, S., SALVATI, N. and SCHMID, T. (2023). A framework for producing small area estimates based on area-level models in R. *R J.* **15** 316–341.

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

JIONGO, V. D., HAZIZA, D. and DUCHESNE, P. (2013). Controlling the bias of robust small-area estimators. *Biometrika* **100** 843–858. MR3142336 https://doi.org/10.1093/biomet/ast030

KILIC, T., SERAJUDDIN, U., UEMATSU, H. and YOSHIDA, N. (2017). Costing household surveys for monitoring progress toward ending extreme poverty and boosting shared prosperity. World Bank Policy Research Working Paper 7951.

KRENNMAIR, P. and SCHMID, T. (2022). Flexible domain prediction using mixed effects random forests. *J. R. Stat. Soc.*, *Ser. C*, *Appl. Stat.* **71** 1865–1894. MR4511133 https://doi.org/10.1111/rssc.12600

KRENNMAIR, P., WÜRZ, N. and SCHMID, T. (2026). Analysing opportunity cost of care work using mixed effects random forests under aggregated auxiliary data. *J. R. Stat. Soc.*, *Ser. C*, *Appl. Stat.* **75** 1–20. MR5015066 https://doi.org/10.1093/jrsssc/qlaf031

KRENNMAIR, P., WÜRZ, N., SCHMID, T. and TZAVIDIS, N. (2026). Supplement to "Random forests and mixed effects random forests for small area estimation of general parameters: a poverty mapping case study in Mozambique." https://doi.org/10.1214/25-AOAS2126SUPPA, https://doi.org/10.1214/25-AOAS2126SUPPB

KREUTZMANN, A.-K., PANNIER, S., ROJAS-PERILLA, N., SCHMID, T., TEMPL, M. and TZAVIDIS, N. (2019). The R package emdi for estimating and mapping regionally disaggregated indicators. *J. Stat. Softw.* **91** 1–33.

LEE, Y., ROJAS-PERILLA, N., RUNGE, M. and SCHMID, T. (2023). Variable selection using conditional AIC for linear mixed models with data-driven transformations. *Stat. Comput.* **33** Paper No. 27, 17. MR4526363 https://doi.org/10.1007/s11222-022-10198-9

MARINO, M. F., RANALLI, M. G., SALVATI, N. and ALFÒ, M. (2019). Semiparametric empirical best prediction for small area estimation of unemployment indicators. *Ann. Appl. Stat.* **13** 1166–1197. MR3963567 https://doi.org/10.1214/18-AOAS1226

MENDEZ, G. (2008). Tree-based methods to model dependent data. PhD thesis, Arizona State Univ.

MENDEZ, G. and LOHR, S. (2011). Estimating residual variance in random forest regression. *Comput. Statist. Data Anal.* **55** 2937–2950. MR2813057 https://doi.org/10.1016/j.csda.2011.04.022

MOLINA, I. and RAO, J. N. K. (2010). Small area estimation of poverty indicators. *Canad. J. Statist.* **38** 369–385. MR2730115 https://doi.org/10.1002/cjs.10051

OPSOMER, J. D., CLAESKENS, G., RANALLI, M. G., KAUERMANN, G. and BREIDT, F. J. (2008). Non-parametric small area estimation using penalized spline regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 265–286. MR2412642 https://doi.org/10.1111/j.1467-9868.2007.00635.x

PFEFFERMANN, D. (2013). New important developments in small area estimation. *Statist. Sci.* **28** 40–68. MR3075338 https://doi.org/10.1214/12-STS395

PRASAD, N. G. N. and RAO, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.* **85** 163–171. MR1137362

PROBST, P., WRIGHT, M. and BOULESTEIX, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **9** e1301.

RAO, J. N. K. and MOLINA, I. (2015). *Small Area Estimation*, 2nd ed. *Wiley Series in Survey Methodology*. Wiley, Hoboken, NJ. With a foreword by Graham Kalton. MR3380626 https://doi.org/10.1002/9781118735855

ROJAS-PERILLA, N., PANNIER, S., SCHMID, T. and TZAVIDIS, N. (2020). Data-driven transformations in small area estimation. *J. R. Stat. Soc. A* **183** 121–148. MR4049657 https://doi.org/10.1111/rssa.12488

SEXTON, J. and LAAKE, P. (2009). Standard errors for bagged and random forest estimators. *Comput. Statist. Data Anal.* **53** 801–811. MR2654590 https://doi.org/10.1016/j.csda.2008.08.007

SINGLETON, A., ALEXIOU, A. and SAVANI, R. (2020). Mapping the geodemographics of digital inequality in Great Britain: An integration of machine learning into small area estimation. *Comput. Environ. Urban Syst.* **82** 101486.

SUGASAWA, S. and KUBOKAWA, T. (2017). Transforming response values in small area prediction. *Comput. Statist. Data Anal.* **114** 47–60. MR3660838 https://doi.org/10.1016/j.csda.2017.03.017

SUGASAWA, S. and KUBOKAWA, T. (2019). Adaptively transformed mixed-model prediction of general finite-population parameters. *Scand. J. Stat.* **46** 1025–1046. MR4033802 https://doi.org/10.1111/sjos.12380

TZAVIDIS, N., MARCHETTI, S. and CHAMBERS, R. (2010). Robust estimation of small-area means and quantiles. *Aust. N. Z. J. Stat.* **52** 167–186. MR2723026 https://doi.org/10.1111/j.1467-842X.2010.00572.x

TZAVIDIS, N., ZHANG, L.-C., LUNA, A., SCHMID, T. and ROJAS-PERILLA, N. (2018). From start to finish: A framework for the production of small area official statistics. *J. R. Stat. Soc. A* **181** 927–979. MR3876378 https://doi.org/10.1111/rssa.12364

UNITED NATIONS (2015). Transforming our world: the 2030 Agenda for Sustainable Development. New York, United Nations Resolution A/RES/70/1.

VILJANEN, M., MEIJERINK, L., ZWAKHALS, L. et al. (2022). A machine learning approach to small area estimation: Predicting the health, housing and well-being of the population of Netherlands. *Int. J. Health Geogr.* **21**.

WAGER, S., HASTIE, T. and EFRON, B. (2014). Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. *J. Mach. Learn. Res.* **15** 1625–1651. MR3225243

WELSH, A. H. and RONCHETTI, E. (1998). Bias-calibrated estimation from sample surveys containing outliers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 413–428. MR1616069 https://doi.org/10.1111/1467-9868.00133

WORLD BANK GROUP (2023). Mozambique Poverty Assessment, June 2023: Poverty Reduction Setback in Times of Compounding Shocks License: CC BY 3.0 IGO. Available from http://hdl.handle.net/10986/40106.

WRIGHT, M. N. and ZIEGLER, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77** 1–17.

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 833–855
https://doi.org/10.1214/25-AOAS2135
© Institute of Mathematical Statistics, 2026

SMALL AREA ESTIMATION OF EDUCATION LEVELS IN LOW- AND MIDDLE-INCOME COUNTRIES

BY YUNHAN WU[1,a], AMEER DHARAMSHI[1,b] AND JON WAKEFIELD[2,c]

[1]*Department of Biostatistics, University of Washington,* [a]*yunhanwu@uw.edu,* [b]*adharams@uw.edu*
[2]*Department of Statistics and Department of Biostatistics, University of Washington,* [c]*jonno@uw.edu*

Education is a key driver of social and economic mobility, yet disparities in attainment persist, particularly in low- and middle-income countries (LMICs). Existing indicators, such as mean years of schooling for adults aged 25 and older (MYS25) and expected years of schooling (EYS), offer a snapshot of an educational system, but lack either cohort-specific or temporal granularity. To address these limitations, we introduce the ultimate years of schooling (UYS)—a birth cohort-based metric targeting the final educational attainment of any individual cohort, including those with ongoing schooling trajectories. As with many attainment indicators, we propose to estimate UYS with cross-sectional household surveys. However, for younger cohorts, estimation fails, because these individuals are right-censored leading to severe downward bias. To correct for this, we propose to reframe educational attainment as a time-to-event process and deploy discrete-time survival models that explicitly account for censoring in the observations. At the national level, we estimate the parameters of the model using survey-weighted logistic regression, while for finer spatial resolutions, where sample sizes are smaller, we embed the discrete-time survival model within a Bayesian spatiotemporal framework to improve stability and precision. Applying our proposed methods to data from the 2022 Tanzania Demographic and Health Surveys, we estimate female educational trajectories corrected for censoring biases and reveal substantial subnational disparities. By providing a dynamic, bias-corrected, and spatially disaggregated measure, our approach enhances education monitoring; it equips policymakers and researchers with a more precise tool for monitoring current progress toward education goals and for designing future targeted policy interventions in LMICs.

REFERENCES

ALLISON, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociol. Method.* **13** 61–98.

BARRO, R. J. and LEE, J.-W. (1993). International comparisons of educational attainment. *J. Monet. Econ.* **32** 363–394.

BESAG, J., YORK, J. and MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.* **43** 1–20. MR1105822 https://doi.org/10.1007/BF00116466

BINDER, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Int. Stat. Rev.* **51** 279–292. MR0731144 https://doi.org/10.2307/1402588

BROWN, C. C. (1975). On the use of indicator variables for studying the time-dependence of parameters in a response-time model. *Biometrics* **31** 863–872.

CROFT, T. N., ALLEN, C. K., ZACHARY, B. W. et al. (2023). *Guide to DHS Statistics*. ICF, Rockville, MD, USA.

DELPRATO, M., CHUDGAR, A. and FROLA, A. (2024). Spatial education inequality for attainment indicators in sub-Saharan Africa and spillovers effects. *World Dev.* **176** 106522.

DELPRATO, M. and FROLA, A. (2022). Zones of educational exclusion of out-of-school youth. *Int. J. Educ. Soc. Dev.* **88** 102532.

DHARAMSHI, A., BARAKAT, B., ALKEMA, L. and ANTONINIS, M. (2022). A Bayesian model for estimating Sustainable Development Goal indicator 4.1.2: School completion rates. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **71** 1822–1864. MR4511132 https://doi.org/10.1111/rssc.12595

Key words and phrases. Educational attainment, Bayesian smoothing, spatial models, time series models, censoring, survival modeling, small area estimation.

DONG, T. Q. and WAKEFIELD, J. (2021). Modeling and presentation of vaccination coverage estimates using data from household surveys. *Vaccine* **39** 2584–2594.

FUCHS, R., PAMUK, E. and LUTZ, W. (2010). Education or wealth: Which matters more for reducing child mortality in developing countries? *Vienna Yearb. Popul. Res.* **8** 175–199.

GAO, Y., KENNEDY, L., SIMPSON, D. and GELMAN, A. (2021). Improving multilevel regression and poststratification with structured priors. *Bayesian Anal.* **16** 719–744. MR4303866 https://doi.org/10.1214/20-BA1223

GLOBAL ADMINISTRATIVE AREAS (2022). GADM database of global administrative areas, version 4.1.

GRAETZ, N., ZIMMERMAN, A., BURSTEIN, R., BIEHL, M., SHIELDS, C., MOSSER, J., CASEY, D., DESHPANDE, A., EARL, L. et al. (2018). Mapping local variation in educational attainment across Africa. *Nature* **555** 48–53.

HÁJEK, J. (1971). Discussion of 'an essay on the logical foundations of survey sampling, part I', by D. Basu. *Found. Stat. Inference* **326**.

KALBFLEISCH, J. D. and PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, New York. MR1924807 https://doi.org/10.1002/9781118032985

KNORR-HELD, L. (2000). Bayesian modeling of inseparable space–time variation in disease risk. *Stat. Med.* **19** 2555–2567.

LEWIN, K. M. (2009). Access to education in sub-Saharan Africa: Patterns, problems and possibilities. *Comp. Educ.* **45** 151–174.

LUMLEY, T. (2010). *Complex Surveys*: *A Guide to Analysis Using R*. Wiley, New York.

LUTZ, W., CUARESMA, J. C. and SANDERSON, W. (2008). The demography of educational attainment and economic growth. *Science* **319** 1047–1048.

MACHARIA, P. M., MOTURI, A. K., MUMO, E., GIORGI, E., OKIRO, E. A., SNOW, R. W. and RAY, N. (2023). Modelling geographic access and school catchment areas across public primary schools to support subnational planning in Kenya. *Child. Geogr.* **21** 832–848.

MANDA, S. and MEYER, R. (2005). Age at first marriage in Malawi: A Bayesian multilevel analysis using a discrete time-to-event model. *J. R. Stat. Soc.*, *Ser. A* **168** 439–455. MR2119410 https://doi.org/10.1111/j.1467-985X.2005.00357.x

MEARA, E. R., RICHARDS, S. and CUTLER, D. M. (2008). The gap gets bigger: Changes in mortality and life expectancy, by education, 1981–2000. *Health Aff.* **27** 350–360.

MINISTRY OF EDUCATION, TANZANIA (2017). Education sector development plan (ESDP) 2016/17–2020/21: Delivering quality education and training to all Tanzanians. Retrieved from https://www.globalpartnership.org/sites/default/files/2019-04-gpe-tanzania-esp.pdf.

RIEBLER, A., SØRBYE, S. H., SIMPSON, D. and RUE, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Stat. Methods Med. Res.* **25** 1145–1165. MR3541089 https://doi.org/10.1177/0962280216660421

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. With discussion. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

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

SINGER, J. D. and WILLETT, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *J. Educ. Stat.* **18** 155–195.

SMITS, J. and PERMANYER, I. (2019). The subnational human development database. *Sci. Data* **6** 1–15.

STEVENS, F. R., GAUGHAN, A. E., LINARD, C. and TATEM, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS ONE* **10** 1–22.

SURESH, K., SEVERN, C. and GHOSH, D. (2022). Survival prediction models: An introduction to discrete-time modeling. *BMC Med. Res. Methodol.* **22** 207.

THOMSON, D. R., SHITOLE, S., SHITOLE, T., SAWANT, K., SUBBARAMAN, R., BLOOM, D. E. and PATIL-DESHMUKH, A. (2014). A system for household enumeration and re-identification in densely populated slums to facilitate community research, education, and advocacy. *PLoS ONE* **9** 1–9.

UIS (2025). UIS Stat: SDG 4 Regional Averages.

UNESCO (2024). Global Education Monitoring Report 2024/5. Technical report, UNESCO.

UNESCO INSTITUTE FOR STATISTICS AND AFRICAN UNION (2024). 2024 spotlight on basic education completion and foundational learning in Africa. Technical report, UNESCO.

UNITED NATIONS (2015). Transforming our world: the 2030 agenda for sustainable development. Retrieved from https://sdgs.un.org/2030agenda.

UNITED NATIONS CHILDREN'S FUND (UNICEF) (2019). *Every Child Learns: UNICEF Education Strategy 2019–2030*. UNICEF, New York.

WAKEFIELD, J., FUGLSTAD, G.-A., RIEBLER, A., GODWIN, J., WILSON, K. and CLARK, S. J. (2019). Estimating under-five mortality in space and time in a developing world context. *Stat. Methods Med. Res.* **28** 2614–2634. MR4000184 https://doi.org/10.1177/0962280218767988

WILLETT, J. B. and SINGER, J. D. (1991). From whether to when: New methods for studying student dropout and teacher attrition. *Rev. Educ. Res.* **61** 407–450.

WU, Y., DHARAMSHI, A. and WAKEFIELD, J. (2026). Supplement to "Small Area Estimation of Education Levels in Low- and Middle-Income Countries." https://doi.org/10.1214/25-AOAS2135SUPPA, https://doi.org/10.1214/25-AOAS2135SUPPB

WU, Y. and WAKEFIELD, J. (2024). Modelling urban/rural fractions in low- and middle-income countries. *J. R. Stat. Soc., Ser. A* **187** 811–830. MR4808892 https://doi.org/10.1093/jrsssa/qnae003

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 856–872
https://doi.org/10.1214/26-AOAS2146

A PARTIALLY COLLAPSED GIBBS SAMPLING ALGORITHM FOR REGRESSION WITH MISREPORTED RESPONSE

BY JIAYING WANG[1,a], WEINING SHEN[2,b] AND YUAN WANG[??,c]

[1] *School of Economic Sciences, Washington State University,* [a]*jiaying.wang2@wsu.edu*
[2] *Department of Statistics, University of California, Irvine,* [b]*weinings@uci.edu*
[3] *Department of Mathematics and Statistics, Washington State University,* [c]*yuan.wang@wsu.edu*

In this paper our objective is to identify the risk factors associated with adolescent marijuana use in Washington State, utilizing data from the 2018 and 2021 Healthy Youth Survey (HYS). Despite the survey's assurance of anonymity, the possibility of over- or underreporting exists due to various reasons, such as fear of being exposed, social stigma, and peer pressure. We are also interested in identifying factors that are associated with the occurrence of misreport. To achieve these goals, we develop a full Bayesian framework with a two-level latent linear regression model. The top level is for the true marijuana use response, and the second level is for the occurrence of misreporting. An informative prior is designed to seamlessly incorporate the domain knowledge or prior information while minimizing the risk of prior misspecification. We propose a partially collapsed Gibbs sampling algorithm with a Metropolis–Hastings step to sample the regression coefficients. Simulation studies have been conducted to demonstrate the superior performance of the proposed method over alternative approaches. Our analysis of HYS data discovers multiple factors for identifying at-risk adolescents and informing future prevention efforts.

REFERENCES

SUBSTANCE ABUSE AND MENTAL HEALTH SERVICES ADMINISTRATION (2021). Preventing marijuana use among youth SAMHSA Publication No. PEP21-06-01-001. Technical Report, National Mental Health and Substance Use Policy Laboratory.

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

BALCOMBE, K., BAILEY, A., CHALAK, A. and FRASER, I. (2007). Bayesian estimation of willingness-to-pay where respondents mis-report their preferences. *Oxf. Bull. Econ. Stat.* **69** 413–438.

BALCOMBE, K. and FRASER, I. (2009). Dichotomous-choice contingent valuation with 'dont know' responses and misreporting. *J. Appl. Econometrics* **24** 1137–1152. MR2750282 https://doi.org/10.1002/jae.1109

BOLLINGER, C. R. and DAVID, M. H. (1997). Modeling discrete choice with response error: Food stamp participation. *J. Amer. Statist. Assoc.* **92** 827–835.

BOLLINGER, C. R. and DAVID, M. H. (2001). Estimation with response error and nonresponse: Food-stamp participation in the SIPP. *J. Bus. Econom. Statist.* **19** 129–141. MR1963279 https://doi.org/10.1198/073500101316970368

BOLLINGER, C. R. and VAN HASSELT, M. (2017). Bayesian moment-based inference in a regression model with misclassification error. *J. Econometrics* **200** 282–294. MR3684979 https://doi.org/10.1016/j.jeconom.2017.06.011

BUONACCORSI, J. P. (2010). *Measurement Error: Models, Methods, and Applications. Interdisciplinary Statistics.* CRC Press, Boca Raton, FL. MR2682774 https://doi.org/10.1201/9781420066586

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed. *Monographs on Statistics and Applied Probability* **105**. CRC Press/CRC, Boca Raton, FL. MR2243417 https://doi.org/10.1201/9781420010138

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

Key words and phrases. Bayesian analysis, latent variable, Markov chain Monte Carlo, posterior predictive distribution.

DAVERN, M., KLERMAN, J. A., ZIEGENFUSS, J., LYNCH, V. and GREENBERG, G. (2009). A partially corrected estimate of medicaid enrollment and uninsurance: Results from an imputational model developed off linked survey and administrative data. *J. Econ. Soc. Meas.* **34** 219–240.

DIEBOLT, J. and ROBERT, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *J. Roy. Statist. Soc. Ser. B, Methodol.* **56** 363–375. MR1281940

GREENE, W., HARRIS, M. N., SRIVASTAVA, P. and ZHAO, X. (2018). Misreporting and econometric modelling of zeros in survey data on social bads: An application to cannabis consumption. *Health Econ.* **27** 372–389.

GUSTAFSON, P. (2005). On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables. *Statist. Sci.* **20** 111–140. MR2183445 https://doi.org/10.1214/088342305000000098

GUSTAFSON, P. (2009). What are the limits of posterior distributions arising from nonidentified models and why should we care? *J. Amer. Statist. Assoc.* **104** 1682–1695. MR2750585 https://doi.org/10.1198/jasa.2009.tm08603

GUSTAFSON, P. (2010). Bayesian inference for partially identified models. *Int. J. Biostat.* **6** 17. MR2602560 https://doi.org/10.2202/1557-4679.1206

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

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

HAUSMAN, J. A., ABREVAYA, J. and SCOTT-MORTON, F. M. (1998). Misclassification of the dependent variable in a discrete-response setting. *J. Econometrics* **87** 239–269. MR1649270 https://doi.org/10.1016/S0304-4076(98)00015-3

HEALTHY YOUTH SURVEY (2024). AskHYS. Available at http://www.askhys.net.

JONES, G., JOHNSON, W. O., HANSON, T. E. and CHRISTENSEN, R. (2010). Identifiability of models for multiple diagnostic testing in the absence of a gold standard. *Biometrics* **66** 855–863. MR2758221 https://doi.org/10.1111/j.1541-0420.2009.01330.x

LEWBEL, A. (2000). Identification of the binary choice model with misclassification. *Econometric Theory* **16** 603–609. MR1790293 https://doi.org/10.1017/S0266466600164060

LITTLE, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *J. Amer. Statist. Assoc.* **83** 1198–1202. MR0997603

MCFADDEN, D. L. (1984). Econometric analysis of qualitative response models. *Handb. Econom.* **2** 1395–1457.

MCGLOTHLIN, A., STAMEY, J. D. and SEAMAN, J. W. JR. (2008). Binary regression with misclassified response and covariate subject to measurement error: A Bayesian approach. *Biom. J.* **50** 123–134. MR2414643 https://doi.org/10.1002/bimj.200710402

MEYER, B. D. and MITTAG, N. (2017). Misclassification in binary choice models. *J. Econometrics* **200** 295–311. MR3684980 https://doi.org/10.1016/j.jeconom.2017.06.012

MEYER, B. D., MITTAG, N. and GEORGE, R. M. (2020). Errors in survey reporting and imputation and their effects on estimates of food stamp program participation. *J. Hum. Resour.* 0818–9704R2.

MURPHY, S. M. and ROSENMAN, R. (2019). The "real" number of Washington state adolescents using marijuana, and why: A misclassification analysis. *Subst. Use Misuse* **54** 89–96.

MWALILI, S. M., LESAFFRE, E. and DECLERCK, D. (2005). A Bayesian ordinal logistic regression model to correct for interobserver measurement error in a geographical oral health study. *J. R. Stat. Soc., Ser. C* **54** 77–93. MR2134599 https://doi.org/10.1111/j.1467-9876.2005.00471.x

NARANJO, L., MARTÍN, J., PÉREZ, C. J. and RUFO, M. J. (2014). Addressing misclassification for binary data: Probit and t-link regressions. *J. Stat. Comput. Simul.* **84** 2187–2213. MR3223620 https://doi.org/10.1080/00949655.2013.787424

NARANJO, L., PÉREZ, C. J., MARTÍN, J., MUTSVARI, T. and LESAFFRE, E. (2019). A Bayesian approach for misclassified ordinal response data. *J. Appl. Stat.* **46** 2198–2215. MR3983556 https://doi.org/10.1080/02664763.2019.1582613

WASHINGTON STATE DEPARTMENT OF HEALTH (2022). Washington State healthy youth survey 2021: Analytic Report. Washington State Department of Health.

PARK, T. and VAN DYK, D. A. (2009). Partially collapsed Gibbs samplers: Illustrations and applications. *J. Comput. Graph. Statist.* **18** 283–305. MR2749833 https://doi.org/10.1198/jcgs.2009.08108

PAULINO, C. D., SILVA, G. and ACHCAR, J. A. (2005). Bayesian analysis of correlated misclassified binary data. *Comput. Statist. Data Anal.* **49** 1120–1131. MR2143061 https://doi.org/10.1016/j.csda.2004.07.004

POON, W.-Y. and WANG, H.-B. (2010). Bayesian analysis of multivariate probit models with surrogate outcome data. *Psychometrika* **75** 498–520. MR2719940 https://doi.org/10.1007/s11336-010-9164-6

RAGHUNATHAN, T. E., LEPKOWSKI, J. M., VAN HOEWYK, J., SOLENBERGER, P. et al. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv. Methodol.* **27** 85–96.

TAMER, E. (2010). Partial identification in econometrics. *Ann. Rev. Econ.* **2** 167–195.

TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82** 528–550. MR0898357

TENNEKOON, V. and ROSENMAN, R. (2016). Systematically misclassified binary dependent variables. *Comm. Statist. Theory Methods* **45** 2538–2555. MR3483333 https://doi.org/10.1080/03610926.2014.887105

TOURANGEAU, R. and YAN, T. (2007). Sensitive questions in surveys. *Psychol. Bull.* **133** 859.

VAN BUUREN, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Stat. Methods Med. Res.* **16** 219–242. MR2371007 https://doi.org/10.1177/0962280206074463

VAN DYK, D. A. and PARK, T. (2008). Partially collapsed Gibbs samplers: Theory and methods. *J. Amer. Statist. Assoc.* **103** 790–796. MR2524010 https://doi.org/10.1198/016214508000000409

VAN HASSELT, M., BOLLINGER, C. R. and BRAY, J. W. (2022). A Bayesian approach to account for misclassification in prevalence and trend estimation. *J. Appl. Econometrics* **37** 351–367. MR4400207 https://doi.org/10.1002/jae.2879

WANG, J., SHEN, W. and WANG, Y. (2026). Supplement to "A partially collapsed Gibbs sampling algorithm for regression with misreported response." https://doi.org/10.1214/26-AOAS2146SUPP

The Annals of Applied Statistics
2026, Vol. 20, No. 1, 873–892
https://doi.org/10.1214/25-AOAS2132

NFL GHOSTS: A FRAMEWORK FOR EVALUATING DEFENDER POSITIONING WITH CONDITIONAL DENSITY ESTIMATION

By Ronald Yurko[1,a], Quang Nguyen[1,b] and Konstantinos Pelechrinis[2,c]

[1]*Department of Statistics & Data Science, Carnegie Mellon University,* [a]*ryurko@stat.cmu.edu,* [b]*quang@stat.cmu.edu*
[2]*Department of Informatics and Networked Systems, University of Pittsburgh,* [c]*kpele@pitt.edu*

Player attribution in American football remains an open problem due to the complex nature of 22 players interacting on the field, but the granularity of player tracking data provides ample opportunity for novel approaches. In this work we introduce the first public framework to evaluate spatial and trajectory tracking data of players relative to a baseline distribution of "ghost" defenders. We demonstrate our framework in the context of modeling the nearest defender positioning at the moment of catch. In particular, we provide estimates of how much better or worse their observed positioning and trajectory compared to the expected play value of ghost defenders. Our framework leverages multidimensional tracking data features through flexible random forests for conditional density estimation in two ways: (1) to model the distribution of receiver yards gained enabling the estimation of within-play expected value and (2) to model the 2D spatial distribution of baseline ghost defenders. We present novel metrics for measuring player and team performance based on tracking data, and discuss challenges that remain in extending our framework to other aspects of American football.

REFERENCES

Barnett, V. and Hilditch, S. (1993). The effect of an artificial pitch surface on home team performance in football (soccer). *J. Roy. Statist. Soc., Ser. A, Statist. Soc.* **156**. https://doi.org/10.2307/2982859

Baumer, B. S., Matthews, G. J. and Nguyen, Q. (2023). Big ideas in sports analytics and statistical tools for their investigation. *Wiley Interdiscip. Rev.: Comput. Stat.* **15** e1612. MR4662504 https://doi.org/10.1002/wics.1612

Breiman, L. (2001). Random forests. *Mach. Learn.* **45** 5–32. https://doi.org/10.1023/a:1010933404324

Brill, R. S., Yurko, R. and Wyner, A. J. (2025). Analytics, have some humility: A statistical view of fourth-down decision making. *Amer. Statist.* **79** 393–409. MR4942036 https://doi.org/10.1080/00031305.2025.2475801

Burke, B. (2009). Expected point values. Available at https://www.advancedfootballanalytics.com/2009/12/expected-point-values.html.

Carl, S. (2024). nflplotR: NFL logo plots in 'ggplot2' and 'gt' R package version 1.3.1. Available at https://github.com/nflverse/nflplotR.

Carl, S. and Baldwin, B. (2024). nflfastR: Functions to ffficiently access NFL play by play data R package version 4.6.1. Available at https://CRAN.R-project.org/package=nflfastR.

Carroll, B. N., Palmer, P., Thorn, J. and Pietrusza, D. (1988). *The Hidden Game of Football*. Total Sports, Inc., New York.

Carter, V. and Machol, R. E. (1971). Operations research on football. *Oper. Res.* **19** 541–544. https://doi.org/10.1287/opre.19.2.541

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

Chu, D., Reyers, M., Thomson, J. and Wu, L. Y. (2020). Route identification in the National Football League. *J. Quant. Anal. Sports* **16** 121–132.

Dalmasso, N., Pospisil, T., Lee, A. B., Izbicki, R., Freeman, P. E. and Malz, A. I. (2020). Conditional density estimation tools in python and R with applications to photometric redshifts and likelihood-free cosmological inference. *Astron. Comput.* **30** 100362.

Key words and phrases. Random forests, high-dimensional data, player tracking data, American football.

DESHPANDE, S. K. and EVANS, K. (2020). Expected hypothetical completion probability. *J. Quant. Anal. Sports* **16** 85–94.

DUTTA, R., YURKO, R. and VENTURA, S. L. (2020). Unsupervised methods for identifying pass coverage among defensive backs with NFL player tracking data. *J. Quant. Anal. Sports* **16** 143–161.

FELSEN, P., LUCEY, P. and GANGULY, S. (2018). Where will they go? Predicting fine-grained adversarial multi-agent motion using conditional variational autoencoders. In *Proceedings of the European Conference on Computer Vision* (*ECCV*) 732–747.

FERNÁNDEZ, J., BORNN, L. and CERVONE, D. (2021). A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. *Mach. Learn.* **110** 1389–1427. MR4274679 https://doi.org/10.1007/s10994-021-05989-6

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

GU, C. and DE SILVA, V. (2023). Deep generative multi-agent imitation model as a computational benchmark for evaluating human performance in complex interactive tasks: a case study in football. arXiv preprint. Available at arXiv:2303.13323.

HOOTEN, M. B., JOHNSON, D. S., MCCLINTOCK, B. T. and MORALES, J. M. (2017). Animal movement: Statistical models for telemetry data. CRC Press.

HOWARD, A., REID, J. E., LOPEZ, M., BLISS, T. and CUKIERSKI, W. (2020). *NFL Big Data Bowl* 2021. Available at https://kaggle.com/competitions/nfl-big-data-bowl-2021.

IZBICKI, R. and LEE, A. B. (2017). Converting high-dimensional regression to high-dimensional conditional density estimation. *Electron. J. Stat.* **11** 2800–2831. MR3679910 https://doi.org/10.1214/17-EJS1302

KOVALCHIK, S. A. (2023). Player tracking data in sports. *Annu. Rev. Stat. Appl.* **10** 677–697. MR4567810

LE, H. M., CARR, P., YUE, Y. and LUCEY, P. (2017a). Data-driven ghosting using deep imitation learning. In *Proceedings of the* 2017 *MIT Sloan Sports Analytics Conference*.

LE, H. M., YUE, Y., CARR, P. and LUCEY, P. (2017b). Coordinated multi-agent imitation learning. In *Proceedings of the* 34*th International Conference on Machine Learning* (D. Precup and Y. W. Teh, eds.). *Proceedings of Machine Learning Research* **70** 1995–2003. PMLR.

LINDSEY, G. R. (1963). An investigation of strategies in baseball. *Oper. Res.* **11** 477–501. https://doi.org/10.1287/opre.11.4.477

LOPEZ, M. J. (2020). Bigger data, better questions, and a return to fourth down behavior: An introduction to a special issue on tracking data in the National Football League. *J. Quant. Anal. Sports* **16** 73–79. https://doi.org/10.1515/jqas-2020-0057

LOWE, Z. (2013). Lights, cameras, revolution. Grantland. Available at https://grantland.com/features/the-toronto-raptors-sportvu-cameras-nba-analytical-revolution/. [Accessed 22-June-2024].

MACDONALD, B. (2020). Recreating the game: Using player tracking data to analyze dynamics in basketball and football. *Harv. Data Sci. Rev.* **2**. https://doi.org/10.1162/99608f92.6e25c7ee

NGUYEN, Q., JIANG, R., ELLINGWOOD, M. and YURKO, R. (2025). Fractional tackles: Leveraging player tracking data for within-play tackling evaluation in American football. *Sci. Rep.* **15** 2148.

NGUYEN, Q., YURKO, R. and MATTHEWS, G. J. (2024). Here comes the STRAIN: Analyzing defensive pass rush in American football with player tracking data. *Amer. Statist.* **78** 199–208. MR4734070 https://doi.org/10.1080/00031305.2023.2242442

NFL FOOTBALL OPERATIONS (2024). Big Data Bowl. NFL.com. Available at https://operations.nfl.com/gameday/analytics/big-data-bowl. [Accessed 22-June-2024].

POSPISIL, T. and LEE, A. B. (2018). RFCDE: Random forests for conditional density estimation. arXiv preprint. Available at arXiv:1804.05753.

POSPISIL, T. and LEE, A. B. (2019). (f)RFCDE: Random forests for conditional density estimation and functional data. arXiv preprint. Available at arXiv:1906.07177.

ROMER, D. (2006). Do firms maximize? Evidence from professional football. *J. Polit. Econ.* **114** 340–365. https://doi.org/10.1086/501171

SCHMID, M., BLAUBERGER, P. and LAMES, M. (2021). Simulating defensive trajectories in American football for predicting league average defensive movements. Frontiers in Sports and Active Living 3. https://doi.org/10.3389/fspor.2021.669845

SEIDL, T., CHERUKUMUDI, A., HARTNETT, A., CARR, P. and LUCEY, P. (2018). Bhostgusters: Realtime interactive play sketching with synthesized NBA defenses. In *Proceedings of the* 2018 *MIT Sloan Sports Analytics Conference*.

SICILIA, A., PELECHRINIS, K. and GOLDSBERRY, K. (2019). DeepHoops: Evaluating micro-actions in basketball using deep feature representations of spatio-temporal data. In *Proceedings of the* 25*th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* ACM, New York. https://doi.org/10.1145/3292500.3330719

SRINIVASAN, P., SUBRAMANIAN, R. and KNOTTENBELT, W. (2023). Thinking the GOAT: Imitating tennis styles. In *Proceedings of the* 2023 *MIT Sloan Sports Analytics Conference*.

R CORE TEAM (2024). *R*: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. Available at https://www.R-project.org/.

WOLFSON, J., ADDONA, V. and SCHMICKER, R. (2017). Forecasting the performance of college prospects selected in the National Football League draft. In *Handbook of Statistical Methods and Analyses in Sports*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 137–163. CRC Press, Boca Raton, FL. MR3837234

YURKO, R., MATANO, F., RICHARDSON, L. F., GRANERED, N., POSPISIL, T., PELECHRINIS, K. and VENTURA, S. L. (2020). Going deep: Models for continuous-time within-play valuation of game outcomes in American football with tracking data. *J. Quant. Anal. Sports* **16** 163–182.

YURKO, R., NGUYEN, Q. and PELECHRINIS, K. (2026). Supplement to “NFL ghosts: A framework for evaluating defender positioning with conditional density estimation.” https://doi.org/10.1214/25-AOAS2132SUPPA, https://doi.org/10.1214/25-AOAS2132SUPPB

YURKO, R., VENTURA, S. and HOROWITZ, M. (2019). nflWAR: A reproducible method for offensive player evaluation in football. *J. Quant. Anal. Sports* **15** 163–183. https://doi.org/10.1515/jqas-2018-0010

ZHAN, E., ZHENG, S., YUE, Y., SHA, L. and LUCEY, P. (2019). Generating multi-agent trajectories using programmatic weak supervision. In 7*th International Conference on Learning Representations*, *ICLR*. 2019.