Kexin Xie

Ph.D. Student, Department of Statistics, Virginia Tech
Tel: (540)824-8239 Email: kexinx@vt.edu

EDUCATION

Virginia Tech | Ph.D. in Statistics (GPA: 4/4) 08/2021 - 05/2026 (expected)

- **Advisor**: Prof. Xinwei Deng.
- **Dissertation**: Modeling and variable selection of high-dimensional data for complex systems.

Dalian University of Tech | B.S. in Mathematics (GPA:3.9/4, top 2/125) 09/2016 - 06/2020

- **Thesis**: Extended Compartmental model based on MCMC for the COVID-19 prediction.

RESEARCH INTERESTS

- **Advanced Statistics in Machine Learning and AI**:
  Large Language Models in statistical programming; Uncertainty quantification and digital twin
- **Advanced Statistical Modeling and Analysis**:
  Variable selection for high-dimensional data; Design and analysis of computer and physical experiments
- **Applied Statistics in Healthcare and Engineering**:
  Data center load forecasting; Statistical methods in biomedicine and clinical trial; Health economical analytics and cost-effectiveness analysis

PROFESSIONAL EXPERIENCE

JMP Statistical Discovery 05/2025-08/2025

Research Statistician & Software Developer Cary, NC, U.S.

- Proposed an optimal sparse projection design for systems with treatment cardinality constraints to enable statistical efficiency for modeling and analysis within feasible constraints.
- Introduced a tailored optimal projection criterion that ensures a good space-filling property in subspaces and promotes orthogonality or near-orthogonality among factors.
- Developed an interactive JMP Add-In to efficiently explore the design space with treatment cardinality constraints and implemented the algorithm on JMP software using JMP-script language.

PJM Interconnection (one of the world largest utility companies) 05/2024 - 08/2024

Ph.D. Summer Intern in Advanced Analytics Audubon, PA, U.S.

- Employed cutting-edge machine learning models, including SVM, XGBoost, CNN, LSTM, and Transformers, for hourly load forecasting in PJM's electricity grid based on 7-year hourly dataset.
- Proposed a Two-Stage Bias Correction procedure to reduce the predicted error due to distribution shift and forecasted features, and can be seamlessly integrated into established industrial operational procedures.
- Achieved a 50%+ reduction in forecasting RMSE compared to prior methods, and a 75%+ reduction in bias used in Intermediate-Term and Real-Time Security Constrained Economic Dispatch.

Biocomplexity Institute, University of Virginia 05/2023 - 08/2023

Summer Research Intern Charlottesville, VA, U.S.

- Applied high-resolution agent-based Compartmental model to simulate the spread of measles on a synthetic yet realistic social network of Virginia with approximately 7.6 million nodes and 371.9 million edges.
- Developed a Bayesian Spatial Autoregressive Mix-Effects Model to quantify the effects of increasing Measles, Mumps, and Rubella (MMR) vaccine rate on disease reduction and economic burden.
- Leveraged integrated nested Laplace approximation (INLA) to estimate the spillover effect of MMR vaccine rate from posterior marginal distribution, providing a faster alternative to MCMC.

SELECTED HONORS & AWARDS

Fellowship & Scholarship

- The Raymond H. Myers Award (top 1/18) 05/2022
  Department of Statistics, Virginia Tech
  - *Awarded for outstanding 1st Year Ph.D. student on Linear Model and Design of Experiment.*

Teaching Award

- The Best SAIG Short Course Development 05/2022
  Department of Statistics, Virginia Tech
  - *Awarded for outstanding student accomplishments in short course development.*

Travel Awards

- Quality and Productivity Research Conference Travel Scholarship 06/2025
  Quality and Productivity (QP) Section, American Statistical Association (ASA)
- JMP-P&G Student-Early Career Travel Award 06/2025
  2025 Annual ASA/IMS Spring Research Conference (SRC)
- Fall Technical Conference (FTC) Student Grant Awards 08/2024
  66th Annual Fall Technical Conference

Other Awards

- Outstanding Graduate in Dalian City, by Dalian City Education Bureau 05/2020
- National Scholarship (top 3%), by Ministry of Education 2017 & 2018 & 2019
- First Level Scholarship (top 5%), by Dalian University of Technology 2017 & 2018 & 2019
- Outstanding Student Award, by Dalian University of Technology 12/2017 & 12/2018

PUBLICATION

- Citations: 43 (Google Scholar, as of Sep 2, 2025)
- Authored/ co-authored 14 journal articles accepted and 1 manuscript to be submitted or under preparation.
- Published articles in leading journals in different fields, such as *The Lancet Global Health (IF: 17.98)*, *BMJ Global Health, Technimetrics, Journal of Quality Technology, Quality Engineering and Energy*.

Peer-Reviewed Journal Articles

1. **Xie, K.**, & Deng, X. (2025). Bi-level variable selection of conditional main effects for generalized linear models. *Technometrics (accepted).*
2. Lian, J., **Xie, K.**, … & Deng, X. (2025). A Statistical Approach to Quality Evaluation of AI Mislabel Detection Algorithm. *Quality Engineering (accepted).*
3. Guo, Q., **Xie, K.**, and Deng, X. (2025). Discussion on "Statistical Solutions for Interdisciplinary Problem-Solving", *Quality Engineering, 1-5.*
4. **Xie, K.**, Marathe, A., Deng, X., Ruiz-Castillo, P., Imputiua, S., ... & Rist, C. (2023). Alternative approaches for creating a wealth index: the case of Mozambique. *BMJ Global Health*, *8*(8), e012639.
5. Ruiz-Castillo, P., Imputiua, S., **Xie, K.**, Elobolobo, E., Nicolas, P., Montaña, J., ... & Sacoor, C. (2023). BOHEMIA a cluster randomized trial to assess the impact of an endectocide-based one health approach to malaria in Mozambique: baseline demographics and key malaria indicators. *Malaria Journal*, *22*(1), 1-12.
6. **Xie, K.**, & Hlynka, M. (2019). Forward and Reverse Parking in a Parking Lot. *Applied Mathematical Sciences*, 13(22), 1091-1102.

Journal Articles Under Revision

7. **Xie, K.**, Otuko, R., … & Rist, C. (2025). Trial-based cost-effectiveness analysis of ivermectin mass drug administration for malaria control in Kwale County, Kenya. *The Lancet Infectious Diseases (minor revision).*
8. **Xie, K.**, Marathe, A., Thakur, M., … & Vullikanti. A. (2025). A health and economic evaluation of the spatial spillover effect from measles resurgence. *Scientific Report (minor revision).*
9. **Xie, K.**, Wu, Y., Deng, X. & Giacomoni, A. (2025). The Big Data Era and Electricity Load Forecasting: Key Lessons from Virginia's Data Centers Surge. *Energy (minor revision).*
10. Wang, Y., **Xie, K.**, & Deng, X. (2025). High-dimensional Data and Machine Learning Models. In *Springer Handbook of Reliability (2nd edition) (under review).*
11. Song, X., **Xie, K.**, Lee, L., Chen, R., Clark, J. M., He, H., ... & Hong, Y. (2025). Performance Evaluation of Large Language Models in Statistical Programming. *Journal of Quality Technology (under review).*
12. Assenga, A., Sale, M., **Xie, K.**, … & Rist, C. (2024). Effect of ivermectin mass drug administration on smallholder pig production in Mopeia District, Mozambique. *Journal of Tropical Animal Health and Production (under review)*.
13. Wei, Q., Chan, V., **Xie, K.**, … & Deng, X. (2025). Estimation of Penalized Single Index Models with Exact Shape Constraints. *Chemometrics and Intelligent Laboratory Systems (under review).*
14. Sanz Gutiérrez A., Otukoc, R. **Xie, K.**, … & Chaccour, C. (2025). Spatial exploration of the impact of cattle on malaria transmission using data from a cluster randomized clinical trial. *BMJ Global Health (under review).*

Journal Articles Ready to Submit

15. **Xie, K.**, Lekivetz, R., & Deng, X. (2025). Optimal Sparse Projection Design for Systems with Treatment Cardinality Constraint. *Journal of Quality Technology.*

RESEARCH EXPERIENCE

Variable Selection of Conditional Main Effects for Generalized Linear Model 05/2022 - Present

- Developed a hierarchical adaptive penalization for variable selection in conditional main effects, achieving a 50% + reduction in the false positive rate and doubling the F1 score compared to LASSO.
- Unitized Coordinate Descent and Iteratively Reweighted Least Square (IRLS) algorithms to achieve faster and stable convergence in complex, high-dimensional problems with 10K+ variables.
- Created the C++ integrated R package "GLMCMEnet" with 8 embedded functions to implement this method, facilitating its application across gene association studies like SNP marker selection.

AI-STAT: AI Powered Statistical Analysis and Performance Evaluations 11/2024 - Present

- Developed a benchmark dataset of 207 SAS programming questions and designed a multi-dimensional evaluation framework assessing Code Correctness, Executability, and Output Quality.
- Led a team of more than 10 experts to evaluate LLM-generated SAS code, including conducting pre-workshops to ensure quality assessment standards and organizing collaboration across multiple domains.
- Conducted comparative evaluations of code solutions generated by GPT-3.5, GPT-4.0, and Llama using the benchmark dataset, analyzing performance across the defined metrics.

Broad One Health Endectocide-based Malaria Intervention in Africa (BOHEMIA) 08/2022 – 12/2024

- Collaborated with 6 institutions in the BOHEMIA project to create a complementary malaria control strategy by ivermectin mass drug administration to humans and livestock through 2 clinical trials in Africa.
- Quantified ivermectin's impact on malaria incidence in 1,975 Mozambican and 2,606 Kenyan children using GEE and Cox models, showing an 21% incident rate ratio reduction and 15% hazard ratio reduction.

- Conducted a health economic evaluation and cost-effectiveness analysis using survey data from 626 households and 2,728 individuals to provide data-driven insights for BOHEMIA's strategic decisions.

Wealth Index for Social and Environmental Systems in Low-Income Regions 05/2022 - 03/2023

- Established a wealth index for Mozambique using DHS methodology based on censes data from 26K+ households, effectively capturing poverty and wealth disparities in low-income regions.
- Proposed a feature-selection PCA method to develop a sparse and robust wealth index, with a 0.76 AUC accuracy for the classifying the households into rich or poor based on 2022 international poverty line.

TEACHING EXPERIENCE

STAT 5984 SS: Linear Models for Data Science, Virginia Tech Fall 2025

Guest Lecturer Blacksburg, VA, U.S.

- Taught a graduate-level statistics section (10+ students; 50-minute sessions).
- Developed and delivered two lectures on regularization method (LASSO, adaptive LASSO, elastic net), with accompanying slides, code, and simulation demoes.

Statistical Application & Innovations Group (SAIG), Virginia Tech 05/2022 - 08/2022

Short Course Developer Blacksburg, VA, U.S.

- Mentored 10+ students on statistical analysis, experimental design, and the use of statistical software, focusing on research in pharmaceutical, biotechnology and chemical sectors.
- Designed training courses on bootstrapping methodology to strengthen students' understanding of data analysis with case study on applications to clinical trials and pharmaceutical research.

Department of Statistics, Virginia Tech 08/2021 - 04/2022

Graduate Teaching Assistant Blacksburg, VA, U.S.

- Collaborated with instructors to design the homework content and organized weekly professional recitation workshops to address and address students' questions for 100+ students with diverse backgrounds.
- Delivered tutoring and recitation sessions to improve students' statistics skills with practical insights, including courses of Biological Statistics, Statistics for Engineering Applications and Biometry.

CONFERENCE & WORKSHOP

Invited talk | 2025 INFORMS Annual Meeting, Atlanta, GA 10/2025

- *Optimal Sparse Projection Design for Systems with Treatment Cardinality Constraint.*

Invited talk | Advances in Interdisciplinary Statistics and Combinatorics symposium, Greensboro, NC 06/2025

- *Bi-level Variable Selection and Machine Learning for Conditional Main Effects.*

Invited talk | Quality and Productivity Research Conference (QPRC), Seattle, WA 06/2025

- *Optimal Sparse Projection Design for Systems with Treatment Cardinality Constraint.*

Contributed talk | Spring Research Conference (SRC), New York, NY 06/2025

- *Bi-level Variable Selection and Machine Learning for Conditional Main Effects.*

Visiting Student | IMSI Long Program, University of Chicago, IN 04/2025

- *Uncertainty Quantification and AI for Complex Systems.*

Contributed talk | Fall Technical Conference, Nashville, TN 10/2024

- *Variable Selection of Conditional Main Effects for Generalized Linear Models with Adaptive Hierarchical Penalty.*

Poster | Design and Analysis of Experiments, Blacksburg, VA 05/2024

- *Variable Selection of Conditional Main Effects with Overlapping Group Structure for Generalized Linear Models.*

PROFESSIONAL SERVICE

Referee and reviewer for:

Journal of the American Statistical Association; Journal of Computational and Graphic Statistic; Journal of Quality Technology; The American Statistician; Journal of the Indian Society for Probability and Statistics; *BMJ Global Health*.

Conference Coordinator, Spring Research Conference (SRC)	06/2025
Session Chair, Quality and Productivity Research Conference (QPRC)	06/2025
Session Chair, 75th Anniversary Department of Statistics Conference	12/2024
Poster Coordinator, Design and Analysis of Experiments Conference	05/2024

PROFESSIONAL MEMBERSHIP

Nominated Associate Member, Sigma Xi, The Scientist Research Honor Society	04/2024-Present
Member, Mu Sigma Rho National Honor Society, Virginia Tech Chapter	04/2023-Present
Student Member, ASA Section on Physical and Engineering Sciences	10/2021-Present