Curriculum Vitae of YAN SONG

Personal Information

Statistics Program,
King Abdullah University of Science and Technology,
Thuwal, 23955-6900, Saudi Arabia

+8613051189305/+9660565056305
yan.song@kaust.edu.sa
YanSong996.github.io

Research Interests

Spatio-Temporal Statistics; Subsampling Methods; Nonparametric Statistics; Computational Statistics and HPC

My work primarily focuses on spatio-temporal statistics, particularly in the analysis of large-scale spatio-temporal data and the development of large- and exa-scale climate emulators, with Gaussian processes being key tools. I have also developed subsampling techniques for various data types and statistical models, with an emphasis on nonparametric statistics.

Education

Renmin University of China, Beijing, China

Ph.D., Statistics, 2018 – 2023

- Advisor: Dr. Wenlin Dai, Ph.D.

Beijing Institute of Technology, Beijing, China

B.S., Statistics, 2014 – 2018

Work Experience

King Abdullah University of Science and Technology, Jeddah, Saudi Arabia

Postdoctoral Fellow, Statistics Program, 13 August 2023 –

- Advisor: Prof. Marc G. Genton, Ph.D.

King Abdullah University of Science and Technology, Jeddah, Saudi Arabia

Visiting student, Statistics Program, 8 March 2022 – 8 February 2023

- Advisor: Prof. Marc G. Genton, Ph.D.

Hong Kong Baptist University, Hong Kong, China

Research assistant, Department of Mathematics, 15 December 2021 – 14 February 2022

- Advisor: Prof. Tiejun Tong, Ph.D.

Honors and Awards

- Gordon Bell Prize for Climate Modelling fanalist for exascale climate emulator at SC24, 2024
- The second prize of Outstanding Papers, National Forum for Doctoral Students in Statistics, 2020
- Outstanding Poster, RUC Youth Forum on Statistics and Data Science, 2020

Publications

1. **Yan Song**, Wenlin Dai, and Marc G. Genton (2024), "Large-scale low-rank Gaussian process prediction with support points," *Journal of the American Statistical Association, Theory and Methods*, to appear, arXiv: 2207.12804.
2. Sameh Abdulah, Allison H. Baker, George Bosilca, Qinglei Cao, Stefano Castruccio, Marc G. Genton, David E. Keyes, Zubair Khalid, Hatem Ltaief, **Yan Song**, Georgiy L. Stenchikov, and Ying Sun (2024), "Boosting earth system model outputs and saving petabytes in their storage using exascale climate emulators," *ACM Gordon Bell Prize for Climate Modelling finalist, International Journal of High Performance Computing Applications*, to appear, arXiv: 2408.04440.
3. **Yan Song**, Zubair Khalid, and Marc G. Genton (2024), "Efficient stochastic generators with spherical harmonic transformation for high-resolution global climate simulations from CESM2-LENS2," *Journal of the American Statistical Association, Applications and Case Studies*, published online, DOI: 10.1080/01621459.2024.2360666.

4. Maoyu Zhang, **Yan Song**, and Wenlin Dai (2024), "Fast robust location and scatter estimation: a depth-based method," *Technometrics*, 66, 14–27, DOI: 10.1080/00401706.2023.2216246.

5. **Yan Song** and Wenlin Dai (2024), "Deterministic subsampling for logistic regression with massive data," *Computational Statistics*, 39, 707–732, DOI: 10.1007/s00180-022-01319-z.

6. Xiaoyu Liu, **Yan Song**, and Kun Zhang (2024), "An exact bootstrap-based bandwidth selection rule for kernel quantile estimators," *Communications in Statistics - Simulation and Computation*, 53, 3699–3720, DOI: 10.1080/03610918.2022.2110595.

7. Yiping Hong, **Yan Song**, Sameh Abdulah, Ying Sun, Hatem Ltaief, David E. Keyes, and Marc G. Genton (2023), "The third competition on spatial statistics for large datasets," *Journal of Agricultural, Biological and Environmental Statistics*, 28, 618–635, DOI: 10.1007/s13253-023-00584-9.

8. Wenlin Dai, **Yan Song(co-first)**, and Dianpeng Wang (2023), "A subsampling method for regression problems based on minimum energy criterion," *Technometrics*, 65, 192–205, DOI: 10.1080/00401706.2022.2127915.

In preparation

1. Yan Song, Zubair Khalid, and Marc G. Genton, "Online regional stochastic generators with Slepian concentration for high-resolution bivariate climate simulations from ERA5".

2. Yan Song, Paolo Giani, Stefano Castruccio, and Marc G. Genton, "Forecasting spatio-temporal wind speed with the combination of multivariate PDE and echo state networks".

3. Yan Song, Wanfang Chen, David Bolin, and Marc G. Genton, "Robust probabilistic forecasting".

Presentations

1. KAUST AMCS-STAT Graduate Seminar, Thuwal, Makkah, Saudi Arabia, 2024 – *Efficient stochastic generators with spherical harmonic transformation for high-resolution global climate simulations from CESM2-LENS2.*

2. Joint Statistical Meetings, Portland, Oregon, USA, 2024 – *Efficient stochastic generators with spherical harmonic transformation for high-resolution global climate simulations from CESM2-LENS2.*

3. KAUST Statistics workshop, Thuwal, Makkah, Saudi Arabia, 2023 – *Efficient stochastic generators with spherical harmonic transformation for high-resolution global climate simulations from CESM2-LENS2.*

4. KAUST Statistics workshop, Thuwal, Makkah, Saudi Arabia, 2022 – *Large-scale low-rank Gaussian process prediction with support points.*

5. National Forum for Doctoral Students in Statistics, Guangzhou, Guangdong, China, 2020 – *A model-free subsampling method based on minimum energy criterion.*

6. RUC Youth Forum on Statistics and Data Science, Beijing, China, 2020 – *A model-free subsampling method based on minimum energy criterion.*

Teaching experiences

1. Half part of short course "Large-Scale Spatial Data Science" at JSM, 2024.
2. A part of short course for the Applied Mathematics and Computational Science and Statistics (AMCS-STAT) school, 2024.
3. One lesson of course STAT 330: Multivariate Statistics at KAUST, 2024.
4. Teaching assistant of course Spatial Statistics at RUC, 2023.
5. Teaching assistant of course Asymptotic Statistics at RUC, 2021 and 2022.
6. Teaching assistant of course Statistical Learning at RUC, 2021.
7. Teaching assistant of course Stochastic Process at RUC, 2020.

Skills and capability

Courses

Probability Theory for Data Science, Statistical Models and Inference, Asymptotic Statistics, Stochastic Process for Data Science, Bayesian Modeling and Inference, Computational Skills for Data Science, Advanced Statistical Computation, Nonparametric Function Estimation, Advanced Applied Statistics and Data Analysis, Statistical Learning.

Operating systems

Mac OS, Microsoft Windows, and Linux

Programming languages

R, Matlab, and Python

GitHub

Example R code can be found at my GitHub profile

Languages

Chinese (mother tongue), English (proficient)

Scholarship and fellowship

- $24,000, Chinese Government Scholarship, China Scholarship Council, 2022
- $1,000, Postgraduate Scientific Research Foundation, Renmin University of China, 2020
- $2,000, Academic Excellence Scholarship, Renmin University of China, 2019, 2020, 2021, 2022, and 2023
- $1,000, Northern Industries Scholarship, Beijing Institute of Technology, 2017

References

- Al-Khawarizmi Distinguished Professor Marc G. Genton
  Statistics Program
  King Abdullah University of Science and Technology (KAUST)
  E-mail: marc.genton@kaust.edu.sa
- Associate Professor Wenlin Dai
  Institute of Statistics and Big Data
  Renmin University of China (RUC)
  E-mail: wenlin.dai@ruc.edu.cn
- Associate Professor Zubair Khalid
  Department of Electrical Engineering
  Lahore University of Management Sciences (LUMS)
  E-mail: zubair.khalid@lums.edu.pk