Huang Huang

Research Scientist
King Abdullah University of Science and Technology

(+966)57-021-3662
huang.huang@kaust.edu.sa
https://hhuang90.github.io

---------- Working Experience

2019-now 

Research Scientist, King Abdullah University of Science and Technology, Saudi Arabia.

- Research modeling of multivariate spatio-temporal data with large size
- Apply machine learning models to environmental data with comparisons to spatio-temporal statistical models
- Lead reading groups on spatial machine learning
- Collaborate with computer scientists in developing high-performance packages for large geostatistical inference
- Mentor students on their internship on research

2018-2019 

Postdoc Fellow II, Advanced Study Program, National Center for Atmospheric Research, U.S.

- Research implementation and optimization of distributed parallel multi-resolution approximation of Gaussian process for large datasets

2017-2018 

Postdoc Fellow, Department of Statistical Science, Duke University & Statistical and Applied Mathematical Sciences Institute, U.S.

- Research statistical methods for climate problems and Bayesian hierarchical modeling of large ensembles

---------- Education

2014-2017 

Ph.D. in Statistics, King Abdullah University of Science and Technology, Saudi Arabia

- Advisor: Prof. Ying Sun
- Thesis title: computational methods for large spatio-temporal datasets and functional data ranking

2011-2014 

Master in Computational Mathematics, Fudan University, China

- Advisor: Prof. Weiguo Gao
- Thesis title: 有向图的随机采样谱稀疏化方法 (Spectral Sparsification Methods of Directed Graphs via Random Sampling, in Chinese)

2007-2011 

Bachelor in Mathematics, Fudan University, China

---------- Honors and Awards

Scholarship

- National scholarship, Fudan University 2012
- First-class graduate scholarship, Fudan University 2012
- Renmin undergraduate scholarship, Fudan University 2011/2010/2009

Travel awards

- Forecasting from Complexity workshop, IMA, U.S. 2018
- Summer school on optimization, SAMSI, U.S. 2016
- Rossbypalooza workshop on climate science and statistics, University of Chicago, U.S. 2016
- Geospatial week by International Society for Photogrammetry and Remote Sensing, France 2015

Poster awards

- Jury's choice second best poster award in Biennial Conference of the Research Group for Environmental Statistics, Italy, by the International Environmetrics Society (TIES) 2015

Others

- Outstanding graduate of Fudan University, China 2014
- First award of National Olympiad in Informatics in Provinces, China 2006

---------- Computational skills

Proficient:	C/C++, Python, R
Intermediate:	Linux, MPI, OpenMP
Basic:	SQL

---------- Publications

Huang, H., Hammerling, D., Li, B., and Smith R. (2021+), Combining Multiple Interdependent Climate Models: A Bayesian Approach, under review.

Huang, H., Genton, M.G., and Sun, Y. (2021+), Visualization of Covariance Structures from Multivariate Spatio-Temporal Random Fields, under review.

Huang, H., Castruccio, S., and Genton, M.G.(2021+), Forecasting High-Frequency Spatio-Temporal Wind Power with Dimensionally Reduced Echo State Networks, *Journal of the Royal Statistical Society Series C (Applied Statistics)*, in revision.

Huang, H., Abdulah, S., Sun, Y., Ltaief, H., Keyes, D. E., and Genton, M. G. (2021), Competition on spatial statistics for large datasets (with discussion), *Journal of Agricultural, Biological, and Environmental Statistics*, to appear.

Blake, L., **Huang H.**, Vanderwende B., Hammerling D. (2021), The Deep-Tree Approach: An Improved Parallel Matlab Implementation of the Multi-resolution Approximation for Massive Spatial Data on High-Performance Computing Systems, *NCAR Technical Note* (NCAR/TN-565-STR).

Salvaña, M.L., Abdulah, S. **Huang, H.**, Ltaief, H., Sun, Y., Genton, M.G., and Keyes, D.E. (2021), High Performance Multivariate Geospatial Statistics on Manycore Systems, *IEEE Transactions on Parallel and Distributed Systems 32*, 2719-2733.

Blake, L., **Huang, H.**, Vanderwende, B., and Hammerling, D. (2019), A Shallow-Tree Multi-resolution Approximation for Distributed and High-Performance Computing Systems, *NCAR Technical Note* (NCAR/TN-559+STR).

Huang, H., Blake, L., and Hammerling, D. (2019), Pushing the Limit: A Hybrid Parallel Implementation of the Multi-resolution Approximation for Massive Data, *NCAR Technical Note* (NCAR/TN-558-STR).

Huang, H. and Sun, Y. (2019), A Decomposition of Total Variation Depth for Understanding Functional Outliers, *Technometrics 61*(4), 445-458.

Huang, H. and Sun, Y. (2018), Hierarchical Low Rank Approximation of Likelihoods for Large Spatial Datasets, *Journal of Computational and Graphical Statistics 27*(1), 110-118.

Huang, H. and Sun Y. (2017), Visualization and Assessment of Spatio-temporal Covariance Properties, *Spatial Statistics 34*, 100272,

Toye, H., Zhan, P., Gapalakrishnan, G., Kartadikaria, R. A., **Huang, H.**, Knio, O., and Hoteit, I. (2017), Ensemble Data Assimilation in the Red Sea: Sensitivity to Ensemble Selection and Atmospheric Forcing, *Ocean Dynamics 67*(7), 915-933.

---------- Presentations and Posters

A Hybrid Parallel Framework of the Multi-Resolution Approximation for Massive Spatial Data

- SIAM Conference on Computational Science and Engineering, online (invited talk) 2021

Functional data depth and its application in the visualization of spatio-temporal covariance structures

- CEMSE Seminar, KAUST (talk) 2020

How is statistics used in geoscience

- Xiamen University, China (invited talk) 2019
- NCAR ASP Seminar, Boulder, U.S. (talk) 2019

Visualization and assessment for properties of spatio-temporal covariance properties

- Forecasting from Complexity, Minneapolis, U.S. (poster) 2018

Total variation depth for functional data

- INFORMS Annual Meeting 2019, Seattle, U.S. (invited talk by *Technometrics*) 2019
- International Conference of the ERCIM WG on Computational and Methodological Statistics, Pisa, Italy. (invited talk) 2018
- Joint Statistical Meetings, Chicago, U.S. (talk) 2016

Inference on the future state of the climate through combining multiple interdependent climate model outputs with observations using Bayesian hierarchical models

- Symposium on Data Science and Statistics, Reston, U.S. (talk) 2018
- Joint Statistical Meetings, Vancouver, Canada (talk) 2018

Hierarchical low rank approximation of likelihoods for large spatial datasets

- Joint Statistical Meetings, Seattle, U.S. (talk) 2015
- International Workshop on Climate Informatics, Boulder, U.S. (poster) 2015
- Spatial Statistics, Avignon, France (poster) 2015
- Biennial Conference of the Research Group for Environmental Statistics, Bari, Italy (poster) 2015

---------- Research Experiences

Bayesian modeling

- Proposal of a Bayesian hierarchical model to infer the future climate states from the interdependent climate models and reanalysis data. 2018
- Fast computations in Bayesian nonparametric regression models, where we apply suitable likelihood approximation techniques. 2017

Computational methods for large datasets

- Fast kriging for large spatial datasets. The proposed hierarchical low rank approximation method is used to do fast spatial interpolation. 2016
- Hierarchical low rank approximation of likelihoods for large spatial datasets. An approximation scheme is proposed to compute the Gaussian likelihood when the covariance matrix is large, dense, and unstructured. 2015

Data assimilation

- Ensemble data assimilation in the Red Sea. An ensemble data assimilation and forecasting system for the Red Sea capable of studying the sensitivity of the system to various filtering parameters and atmospheric forcing is built. 2016

Functional data analysis

- Total variation depth for functional data. A notion of functional data depth is developed for functional data ranking and outlier detection. 2017

High-performance computing

- Investigation of the high-performance multivariate spatial modeling for geostatistical data on manycore systems using the developed package “ExaGeoStat”. 2020
- Implementation and optimization of distributed parallel multi-resolution approximation of Gaussian process for extremely large spatial datasets comprising up to tens of millions of observations using C++. 2019

Industry projects

- Click-through rate prediction. Prediction methods have been developed using massive datasets of user historical behaviors on the distributed file system, Apache Hadoop. 2013
- Data mining in recommendation systems. The Latent Dirichlet Allocation model is used to classify advertisement passages by the 2012

hidden topics, and advertisements are recommended to users accordingly.

Machine learning

- Imputations for biochemical measurements in Argo data. We apply neural network methods to Argo profiles to predict oxygen at locations where the observations are missing. 2018

Spatio-temporal statistics

- Organized the "2021 KAUST Competition on Spatial Statistics for Large Datasets", which attracted 29 research teams worldwide to participate. The competition assesses various state-of-the-art approximation methods for geospatial inference and prediction by using our simulated large synthetic data. 2021
- Formally define different properties of multivariate spatio-temporal covariances and examine them by proposed test functions. 2020
- Visualization and assessment for properties of spatio-temporal covariances, including separability and symmetry, using functional data analysis. 2016

---------- Academia Service

Session Chair Joint Statistical Meetings, Chicago/Seattle, U.S. 2016/2015

Peer-review service

Biometrics, Computational Statistics and Data Analysis, Journal of Agricultural, Biological, and Environmental Statistics, Journal of Climate, Journal of Computational and Graphical Statistics, Stat, Statistics and Computing, Stochastic Environmental Research and Risk Assessment, 4th International Conference on Big Data and Information Analytics

---------- Teaching Experience

International Statistical Institute (ISI) short courses programme, online, 2021 2021

- Short course: Large-scale Spatial Data Science

AMCS-STAT winter school, King Abdullah University of Science and Technology, Saudi Arabia 2021

- Short course: Large-scale Spatial Data Science

Undergraduate Modeling Workshop, North Carolina State University, U.S. 2018

- Short course: R tutorial
- Leading an undergraduate group working on the project, estimation of above ground biomass in the Bonanza Creek experimental forest

Statistics in the Criminal Justice System Workshop, North Carolina Central University, U.S. 2018

- Short course: Hands-on data experience with R

Undergraduate Climate Extremes workshop, SAMSI, U.S. 2017

- Short course: R tutorial

Graduate teaching assistant, KAUST, Saudi Arabia 2016

- Course: Applied statistics and data analysis

Undergraduate teaching assistant, Fudan University, China 2012

- Course: Advanced mathematics

---------- Mentoring Experience

Supervise one undergraduate student from King Fahd University of Petroleum and Minerals for his summer internship on machine learning modeling of solar energy 2021

Supervise one master student from King Abdullah University of Science and Technology for his directed research on spatial and spatio-temporal modeling of COVID-19. 2020

Supervise two undergraduate students from University of Illinois at Urbana-Champaign and Massachusetts Institute of Technology for their summer internships on “Anomaly detection in crowd scenes”	2020
Supervise one master student from King Abdullah University of Science and Technology for his summer research on estimation of reproduction number in a pandemic.	2020
Supervise one Ph.D. student from School of Mines, Colorado, for his summer research on distributed geostatistical modeling	2019