**Mabel Yao**

• mabel.qq.yao@gmail.com • LinkedIn • Info

## EDUCATION

**North Dakota State University** *2026*
PhD, Statistics
focusing on Mathematical Statistics, Machine Learning

**North Dakota State University** *2022*
Master, Computer Science
focusing on Machine Learning, Data Mining

**Tohoku University** *2016*
Master in Engineering, Structural Engineering

**Dalian Jiaotong University** *2013*
Bachelor in Engineering, Civil Engineering & Software Engineering

## Research

- **Interests**

- Mathematical Statistics, Statistical Inference, Machine Learning, Data Science

- **Research Projects by Topics**

* ***Statistical Inference on Graphs*** *2022-present*

- Statistical Methods in Network Analysis

- Graph Representation/ Embedding: Learned and Non-Learned

* ***Algorithmic and Data Modeling, Theoretical Statistics, Machine Learning*** *2022-present*

- Using statistical methods to understand when and why machine learning algorithms work.

- Compare models of Parametric, Semi-parametric, Non-parametric, also consider algorithmic models and data models, to understand how to do model selection for real world applications.

* ***Applications in Tech/IT, Econometric, Financial Economics, Medical Data Science, Drug Discovery/Precision Medicine, etc*** *2022-present*

- Explore applications of statistical methods in real world scenarios, model implementation, evaluation, selection, and validation

- Some use cases are implemented through course projects

- Application of Graphs or Networks in Financial Markets/ Economics, etc.

- Outlier Analysis/ Anomaly Detection in Financial Fraud

- Feature Selection, Analysis and Modeling on Diverse Data: multivariate, time series, survival analysis, etc.

* ***Graph Representation Learning*** *2019-2022*

- Node embedding, graph-level representation learning
- Graph-level representation learning for chemical screening for catalyst discovery
- Molecular property prediction by descriptor based statistical modeling, sequential modeling, graph modeling

* ***Recommendation Systems*** *2019-2022*

- Recommendation in e-commerce and healthcare
- Graph Learning & graph neural networks for Recommendation Systems
- Statistical methods for Recommendations

## PUBLICATION

- **Peer Reviewed Papers**

- Qianqian Yao, Mingao Yuan. Empirical Likelihood test for common invariant subspace of multilayer networks. **(to be finished.)** (2025)
- Mingao Yuan, Qianqian Yao. Testing common degree-correction parameters of multilayer networks. Statistics and Computing (2025) 35:38. https://doi.org/10.1007/s11222-025-10576-z
- Qianqian Yao, Yoshihiro Ito, Yusuke Suzuki, Maeda Masaki. (2015). "Effect of Earthquake Response Spectrum Characteristics on Residual Seismic Performance", Japan Concrete Institute, vol.37 NO.2, 685-690.

- **Conference**

* **2024 Red River Valley Statistics Conference** *Fargo, ND*

Presentations *May, 2024*

- Presentation on time series: Analysis of Time Series Models for Electricity Consumption Forecasting
- Presentation on survival analysis: Survival Analysis for Colorectal Cancer Considering Effects of Staging and Treatment Variables

* **2023 Red River Valley Statistics Conference** *Fargo, ND*

Presentation *May, 2023*

- Presentation on biostatistics: Statistical Learning for Virtual Screening in Drug Discovery

* **2015 Architectural Institute of Japan Annual Conference (Kanto)** *Hiratsuka, JP*

- Linfei Hao, Qianqian Yao, Yusuke Suzuki, Masaki Maeda, "Comparison Between Residual Seismic Capacity Evaluation Method Based on Seismic Capacity Ratio and Recommended by Current Guideline", published in the Architectural Institute of Japan, page 403-404, 2015.9.

* **2014 Architectural Institute of Japan Annual Conference (Kinki)** *Kobe, JP*

- Qianqian Yao, Linfei Hao, Yoshihiro Ito, Yusuke Suzuki, Maeda Masaki, "Seismic Evaluation of Damaged RC buildings Basing on Earthquake Response Spectrum, part 1 Basic Evaluation Basing on Seismic Capacity Index and Effect of Building Capacity Reduction", published in the Architectural Institute of Japan meeting in Koube, page195-196, 2014.8.

- Linfei Hao, Qianqian Yao, Yoshihiro Ito, Yusuke Suzuki, Masaki Maeda, "Residual Seismic Capacity Assessment of Damaged RC Buildings Based on Response Spectrum Part 2. Assessment of residual seismic capacity ratio based on ideal model and comparison with seismic capacity reduction factor", published in the Architectural Institute of Japan, page 197-198, 2014.8.

- **Thesis**

- Qianqian Yao: topic about statistical inference on graphs/ learning in networks **(to be finished.)** (2026). (PhD in Statistics, North Dakota State University)

- Qianqian Yao: Comparison of non-learned and learned molecule representations for catalyst discovery (2022). (Master's in Computer Science, North Dakota State University)

- Qianqian Yao: Residual Seismic Capacity Evaluation for Reinforced Concrete Buildings Considering Effects of Characteristics of Structure and Earthquake Response Spectrum (2016). (Master's in Structural Engineering, Tohoku University)

## Working Experience

- **Teaching Assistantship in North Dakota State University** *Fargo, US*

  ***Department of Statistics*** *2022-present*

  duties: help with computer labs, tutoring, grading, review sessions

  STAT 725 Applied Statistics

  STAT 726 Applied Regression and Analysis of Variance

  STAT 330 Introductory Statistics

  ***Computer Science Department*** *2019-2022*

  duties: help with computer labs, tutoring, grading, review sessions

  CSCI 160 Computer Science I

  CSCI 161 Computer Science II

  CSCI 213 Modern Software Development

  CSCI 114 Computer Applications

  CSCI 122 Visual BASIC

- **Lab Assistant in North Dakota State University** *Fargo, US*

  ***Plant Sciences Department*** *2019-2023*

  Implementation of Experimental Design, including seeding, planting, harvesting, data collection and entry

  Data Analysis

- **Industry Working Experience**

  *Shenzhen Yuanlizhu Engineering Consultants Co.,Ltd* *Shenzhen, China*

  *Structural Engineer* *2017-2019*

Using computer aided engineering tools to design and analyze building structures.

Communicate with clients including investors, constructors, designers to optimize the structural design.

*Shanghai Saiyo Construction Technology Co.,Ltd* *Shanghai, China*

*Project Assistant* *2016-2017*

Participated in a Japanese project of Shopping Mall Construction in Ningbo, and applied Building Information Modeling (BIM) to construct a virtual model of the building for design and clash detection.

*Yamashita Sekkei INC. Tohoku Branch* *Sendai, Japan*

*Intern* *9/2015-10/2015*

Analyze structures with SNAP, created building model, considered seismic isolators and seismic control devices, analyzed seismic response controlled structure and seismic isolation structure to get seismic performance, created animation; Drew construction drawings with AutoCAD.

## Technical Skills

- **Programming Languages**

  R, Python, Julia, SAS, Java, etc. (able to get hands on new languages very quick)

- **Computing and Editing Softwares**

  Latex, RStudio, Anaconda Environment, Matlab, Octave, Minitab, JMP, Visual Studio, etc. (able to get hands on new tools very quick)

- **Domain Knowledge**

  Experienced and Gained Comprehensive Knowledge in Interdisciplinary Applied Sciences and Engineering

## COURSE

- **Related Courses from North Dakota State University** *in-person*
- *STAT886 Advanced Inference*
- *STAT851 Bayesian Statistical Inference*
- *STAT860 Statistical Machine Learning*

  Implementation: Statistical Methods for Recommender System, python

- *STAT874 Generalized Linear Models*

  Implementation: Generalized Linear Models for credit card expenditure analysis by considering variance functions

- *MATH650 Real Analysis I*

  Presentation: Proof of Weierstrass Approximation Theorem

- *STAT672 Time Series*

Implementation: Analysis of Time Series Models for Electricity Consumption Forecasting, R

- *STAT770 Survival Analysis*

  Implementation: Regression Models for Survival Analysis in CRC Considering Staging Groups, R
- *STAT768 Mathematical Statistics II*
- *STAT767 Mathematical Statistics I*
- *STAT764 Multivariate Methods*

  Implementation: Multivariate Analysis for Discrimination of Carcinogenesis Staging, SAS
- *STAT661 Applied Linear Models*

  Implementation: Detection and Evaluation of Outliers by Linear Models, R
- *STAT669 Introduction to Biostatistics*

  Implementation: Descriptor based multiple linear regression model for molecule property prediction, python
- *STAT662 Introduction to Experimental Design*
- *STAT663 Nonparametric Statistics*
- *STAT726 Applied Regression and Variance Analysis*
- *STAT725 Applied Statistics*
- *CSCI 717 Software Construction*

  Implementation: Natural Language Processing: text classification, python
- *CE793 Machine Learning for Engineers*

  Implementation: Multi-label classification based on image similarity, python
- *CSCI846 Distributed Systems*

  Implementation: Distributed database built on client-server architecture, java
- *CSCI679 Introduction to Data Mining*

  Implementation of recommender system based on different models, python
- *CSCI736 Advanced Intelligent Systems*

  Implementation of expert system for real estate recommnendation by drools, java
- *CSCI713 Software Development Processes*
- *CSCI741 Algorithm Analysis*
- *CSCI879 Advanced Data Mining*

  Implementation: Network Mining and analysis using deepwalk, line, and node2vec, python
- *CSCI724 Introduction to Artificial Intelligence*

  Implementation: Large scale study of programming languages and code quality in github, python

- *CSCI765 Introduction to Database Systems*

  Implementation: Evaluation of real estate market using deep learning, python

- **Free Open Source: Theories in Mathematical Statistics and Machine Learning** *online*

  *NPTEL Probability Foundations, Dr. Krishna Jagannathan* *Syllabus*

  *CMU 36-708 Statistical Methods for Machine Learning, Dr. Larry Wasserman* *Syllabus*

  *CMU 36-709 Advanced Statistical Theory I, Dr. Sivaraman Balakrishnan* *Syllabus*

  *CMU 36-710 Advanced Statistical Theory II, Dr.Alessandro Rinaldo* *Syllabus*

  *UCLA STATS 100C Linear Models, Dr. Arash A. Amini* *Syllabus*

  *UCLA STATS 200B Theoretical Statistics, Dr. Arash A. Amini* *Syllabus*

  *UCLA STATS 200C High-dimensional Statistics, Dr. Arash A. Amini* *Syllabus*

  *UCLA STATS 231C Theories of Machine Learning, Dr. Arash A. Amini* *Syllabus*

  *UCSD MATH 181A Mathematical Statistics, Dr. David Quarfoot* *Podcast*

  *UCSD MATH 181B Mathematical Statistics, Dr. David Quarfoot* *Podcast*

  *UCSD MATH 180A Introduction to Probability for Data Science, Dr. Todd Kemp* *Podcast*

  *UManchester MATH38161 Multivariate and Machine Learning, Dr. Korbinian Strimmer* *Podcast*

  *UCLA STATS 100C Linear Models, Dr. Arash A. Amini* *Syllabus*

  *MIT OpenCourseWare: Topics In Mathematics With Applications In Finance* *Syllabus*

  *DavidsonX: Drug Discovery & Medicinal Chemistry, edX* *Syllabus*

- **Verified Courses with Certificates on Coursera** *online*

  *IBM Data Science Specialization* *May, 2023*

  *Machine Learning* *May, 2023*

  *Deep Learning Specialization* *March, 2021*