

# Mengling Qiao

Associate Research Scientist, Columbia University

Tel: +1 332-258-2102 | Email: jo642hk@gmail.com | [LinkedIn](#) | [Google Scholar](#)

## CORE COMPETENCIES

- Alternative Data Research • Time-Series Analysis & Forecasting • Applied ML/DL • Spatial-Temporal Modeling & Analysis
- NLP • Data-Driven Research • Applied Generative AI

## EDUCATION

- Wuhan University** | Direct-Entry Ph.D. in *Geographic Information Science* | CN Sept. 2015 - Dec. 2020
- **Awards:** Graduate studies scholarships × 4 years (10%), The Scholarship for Outstanding Freshmen (5%)
- Nanjing Normal University** | B.Sc. in *Geographic Information Science* | CN Sept. 2011 - Jun. 2015
- **Awards:** Talent Scholarship (1%), Outstanding Student Scholarship (2%), National Encouragement Scholarship (1%)


## APPOINTMENTS

- Columbia University in the City of New York** New York, U.S.
- Associate Research Scientist** | *Earth and Environmental Engineering* Jan. 2025 - present
- Postdoctoral Research Scientist** | *Earth and Environmental Engineering* Jan. 2023 - Dec. 2024
- Guest Lecturer** | *Civil Engineering & Engineering Mechanics* | *Data Science Institute* Jan. 2023 - Dec. 2024
- **Secured & Executed Two Multi-Institutional Projects:** Co-led \$600K projects on data-driven disaster response. Defined technical strategy and coordinated execution in multivariate time-series modeling, forecasting, and data engineering, driving project direction and ensuring all major milestones were met.
  - **Comprehensive Knowledge Integration:** Synthesized 900+ academic papers and 400+ reports using AI tools to guide decisions.
  - **Proposal Development & Technical Leadership:** Coordinated 3+ teams to prepare an NSF \$500K proposal on spatiotemporal ML/DL, graph modeling, and transfer learning for smart drainage simulation. (ongoing).

- The Hong Kong University of Science and Technology** Hong Kong, SAR
- Postdoctoral Fellowship** | *Computer Science & Engineering* Jul. 2022 - Jan. 2023
- **Rapid Proposal Development:** Led a team of 5 PhD students to develop an AR/VR research proposal, completed successfully within one week
  - **Visual Analytics and Pipeline Optimization:** Supported decision-support systems and visualization projects; accelerated delivery by 4 months through improved pipeline design.
  - **Awards:** Recipient of the [Research Talent Hub](#) of Hong Kong (HKD 431,264)

- The Chinese University of Hong Kong** Hong Kong, SAR
- Postdoctoral Fellowship** | *Geography and Resource Management* Jun. 2021 - Jul. 2022
- **Executed Project on Risk Modeling:** Led a \$450K project; developed indices and extended models for spatial-temporal risk quantification and prediction; delivered actionable insights for agencies using predictive analytics.
  - **Outreach:** Delivered invited talks at multiple universities to disseminate findings.

## PROJECTS

- Resilience Metrics for High-Frequency Travel Behavior**  National Science Foundation (2023–2024)
- Developed time-series anomaly detection, cross-correlation models, and network-based resilience metrics by integrating multimodal, high-frequency transportation data (3+ million time-series records spanning subway, taxi, rideshare, and bike).
  - Built automated Python workflows, reducing manual processing by 40%.
- Time-Series Modeling of Indicators for Public Safety** United States-Japan Foundation (2023–2024)
- Modeled 15-year city-level time series of 30+ indicators (climate, mobility, census, etc.) using Bayesian models and multiple ML/DL models to quantify city safety. Achieved  $R^2 = 0.85$  citywide, improving prediction accuracy by  $\approx 3\text{--}5\times$  over weaker baselines.
  - Employed advanced feature analysis (Python Pipeline with SHAP model) to rank predictors and improve accuracy.
- Spatial-Temporal Risk Quantification and Prediction** The Chinese University of Hong Kong (2021–2022)
- Designed PCA-weighted time-series indices integrating mobility and demographic features to assess public health risk

across multiple spatial scales. Achieved  $\approx 60\text{--}65\%$  stronger correlation with ground truth compared to traditional index (correlation improved from 0.48 to 0.79).

- Improved prediction accuracy by extending GTWR methods to capture dynamic time-series patterns in disease spread and community response; Outperformed baseline methods:  $R^2$  improved from 0.23 (OLS) to 0.93 (GTWR) in temporal models ( $\approx 300\%$  gain) and from 0.36 to 0.8 in spatial models ( $\approx 120\%$  gain).

#### **Time-Series Forecasting of Real Estate Market Turning Points** *Natural Resource & Planning Bureau (2018–2019)*

- Developed a forecasting model combining a log-periodic power law (LPPL) formulation and a multi-population genetic algorithm with wavelet elimination to identify time-series housing price turning points on 30K+ property transaction records
- Achieving 25% lower residual errors than simulated annealing (e.g., RSS reduced from 415K to 309K).

#### **Social Media Posts and Street View Images Analytics** *Multiple Studies (2018–2025)*

- **Risk and Sentiment Modeling:** Built NLP (Chinese-BERT-wwm) and geospatial models to analyze 10+ million social media posts; identified public responses to heatwaves (F1-score = 0.97); geospatial models achieved enhanced model explanatory power with  $R^2 = 0.87$ , 1.5x higher than standard regression.
- **Street View-Driven Insights:** Developed computational frameworks integrating semantic and geometric features from 166K+ street view imagery and 125K+ building footprints, and human emotions from 1.6+ million social media posts; applied Generalized Additive Mixed Model to capture nonlinear relationships of built environment and human well-being; Identified key thresholds, such as green view optimal at 45%.
- **Economic Segregation Analysis:** Built multilevel measures of economic segregation using large-scale human activities from 15+ million social media, 121K+ customer review comments, and 39K+ property transaction records. The segregation index reveals finer-grained and more realistic economic segregation patterns than traditional methods.

## HIGHLIGHTS

---

- **Team Leadership:** Led applications for 3 grant attempts (1 funded, 1 under review); As PI for 1 research project; Led a capstone project ranked 1<sup>st</sup> of 6 teams; Supervised 3 graduate-level training projects on NLP, ML/DL, and Graph Theory (delivered 32+ invited classes for 30+ graduates); Mentored 2 graduate students.
- **Cross-Functional Engagement:** Led a 4-hour hands-on workshop on “ML/DL for socio-environmental data analytics” at I-GUIDE Forum 2025.
- **Project Delivery:** Acted as lead representative in a multi-PI research collaboration team, regularly presenting and coordinating across faculty and researchers; Managed multiple concurrent research timelines across universities and international collaborators, consistently delivering output on time.
- **Alternative Data Signals:** Extracted high-frequency signals from everyday life, such as mobility patterns, social sentiment, online discourse, and consumer behavior, providing a complementary perspective beyond conventional datasets and unlocking predictive insights for complex systems.
- **Research Impact**
  - Specialized in long-term and high-frequency time-series modeling across domains
  - Published 15+ peer-reviewed papers in top outlets and 2 book chapters, advancing methods in data-driven time-series analytics and resilient human-environment systems. *Full list:* [Google Scholar](#)
  - Delivered 10+ presentations at top-tier international conferences.
  - Served as Guest Editor for Applied Science Special Issue “Geospatial Data Processing, Mining and Application”.
  - Reviewer of 12+ international journals, delivering 40+ review reports

## SKILLS

---

- **Core Data Science & AI**
  - *Time-Series Analysis & Forecasting:* ARIMA/SARIMA, AR/VAR, Prophet, LSTM/GRU, Bayesian Hierarchical Models, Anomaly Detection
  - *ML & DL:* PyTorch, TensorFlow, Transformers, Scikit-Learn, NumPy, Pandas, GeoPandas. (Random Forest, SVM, CNN, RNN, XGBoost/LightGBM, PCA, CCA, PLS)
  - *Spatial-Temporal Modeling:* TWR, GTWR, Spatio-Temporal Regression, Spatial Panel Data Models.
  - *NLP:* Sentiment Analysis, Topic Modeling, LLMs (BERT, Hugging Face).
  - *Math & Stats:* Probability, Linear Algebra, Statistical Inference.
  - *Graph & Network Science:* Graph-Based Modeling, Centrality Analysis, Modularity Detection, Random Walks.
- **Programming & Tools**
  - *Languages:* Python (10+ yrs), R, SQL, C#, Java
  - *Databases & Platforms:* Jupyter, MongoDB, Oracle
  - *GIS & Visualization:* ArcGIS, QGIS, Gephi, Origin.