**Ph.D. Candidate in Geospatial Information Sciences(GIS), Specialty in Spatial Statistics**

Data Enthusiast and Highly Motivated self-learner with 10+ experience specializing in spatiotemporal data collection, pre-processing, management, imputation, mining, analysis, prediction, and visualization. Seek an opportunity as a Data Scientist.

**CORE COMPETENCY**

* **Programming**: Python (Scikit-Learn, Numpy, Scipy, Statsmodels, PySAL, pandas, NLTK, re, Pystan, Tensorflow, PyTorch, PySpark, etc.), R (Spatstat, Gstat, etc.), SQL, Matlab, SPSS, Excel, AWS (Lambda, etc.), WinBUGS, JavaScript, etc.
* **Statistical Modeling**: Generalized Linear Model, Generalized Linear Mixed Model (GLMM), Logistic Model, Multiple Imputation, Exploratory Spatial Data Analysis (ESDA), Spatial Analysis, Geostatistics, Time-Series Analysis, Dynamic Time Warping (DTW), Bayesian Analysis, Principal Component Analysis (PCA), Factor Analysis (FA), Cross-Decomposition Methods (e.g., Canonical Correlation Analysis (CCA)), Stepwise Regression/LASSO/Ridge/Elastic-Net, segmentation, A/B testing, etc.
* **Machine Learning Algorithms**: Decision Tree, Random Forest, Neural Network Models, Support Vector Machines, etc.
* **Geographical Information Systems**: ArcGIS, ArcGIS Pro, QGIS, Python GIS packages (GDAL, Shapely, etc.), PostGIS
* **Image Processing**: PCI Geomatica, Erdas Imagine, eCognition, ENVI, IDRISI Kilimanjaro, OpenCV, Scikit-Image, etc.
* **Database and Deployment**: PostgreSQL, Amazon Web Service (Dynamodb, S3, RDS, Redshift), Azure (Datalake Store, etc.)
* **Data Visualization Tools**: D3.js, three.js, Tableau, ggplot, matplotlib, CorelDRAW, Ulead PhotoImpact

**WORK EXPERIENCE**

**Graduate Teaching Assistant -** TheUniversity of Texas at Dallas (UT Dallas)September 2015 – Present

* Led 2 team members for 3D visualization and won high appreciation by the National Geospatial-Intelligence Agency in GIS Day.

**National Water Center Research Fellow -** National Oceanic & Atmospheric Administration June 2016 – July 2016

* Formulated a new moisture index and applied filters for detecting flooding extents with 5% and 10% improvements in real-time.

**Graduate Research Associate for NIJ Project -** University of OklahomaApril 2014 – September 2014

* Refined geocoding method by text matching, 20% more match rates, and 80% less time, over the most widely used ESRI geocoder.
* Detected crime patterns by DTW (10TB GPS tracking data) and wrote SQL to extract data from a PostgreSQL database.

**EDUCATION**

**Ph.D. in GIS (Spatial Statistics) UT Dallas, Richardson, TX** August 2015 – May 2020 (expected)

* Convinced three agencies to obtain fine-resolution, multi-sourced datasets (around 1TB big data) for experiments.
* Critiqued the most influential cause of imputation-based measurement error in Econometrics, insufficient to explain biased results.
* Developed experiments to pioneeringly assess impacts of using spatial imputations (e.g., geostatistics) in regression analysis.
* Experimented how spatial sampling strategies (e.g., stratified random sampling) influence imputed values and regression inference.
* Propose imputation algorithms initially to deal with missing data, expected better than the most widely used multiple imputation.

**Master of Arts in Geography, Arizona State University, Tempe, Arizona** August 2011 – December 2013

* Developed segmentation by random walker for plants, superior to the most widely used software, eCognition, saving $5200 a year.

**Master of Arts in Geography, National Taiwan Normal University, Taipei, Taiwan** September 2008–June 2010

* Managed databases for spatiotemporal data from over 400 references across 50 years for predictions with extensive missing data.
* Pioneered predictions of long-term landscape changes by PCA, FA, and ESDA to identify drivers of environmental changes.

**Bachelor of Arts in Geography, National Taiwan Normal University, Taipei, Taiwan** September 2003 – June 2007

**SELF-LEARNING PROJECTS**

**Kaggle Projects** (best prediction performances for providing recommendation of game developments) November 2020 – Present

**Developing Spatial-Temporal Matching Algorithms to Predict Crime Incidents** (large missing data)June 2018 – August 2018

* Predicted 11 crimes (705,874 incidents) by parsing Zillow, Yellow Page, Yelp, and Google Maps with other sources as 484 features.
* Developed component matching algorithms, incorporating cross-decomposition methods, increasing adjusted by 0.5.
* Implemented feature matching algorithms, integrating GLMM with elastic net, to reduce the influences of missing data.
* Designed autocorrelation matching algorithms, extended from Multivariate Adaptive Regression Splines, for lagged incidents.

**Comparisons of Model Specifications to Enhance Bayesian Prediction** (spatial patterns) October 2017 – November 2017

* Adapted Bernardinelli (1995) Spatiotemporal Bayesian model to experiments for evaluating specifications on spatial prediction.

**Multivariate Spatial Crime Analysis** (violations of hypotheses tests in regression) March 2017 – April 2017

* Extended PCA with variance inflation factor to improve modeling performance and prevent violating the multicollinearity test.
* Implemented the Eigenvector Spatial Filtering in Python to handle spatial autocorrelation in residuals.