

KATHERINE KEMPFERT

kempfert@berkeley.edu

EDUCATION

University of California Berkeley

Aug. 2019 - Present

PhD in Statistics

University of Florida

Aug. 2015 - May 2019

Bachelor of Science in Statistics

Summa Cum Laude

Bachelor of Science in Mathematics

Cum Laude

Cumulative GPA: 3.96, Major GPA: 4.0

Honor's Thesis: *Where Does Haydn End and Mozart Begin? Composer Classification of String Quartets*

Advisor: Dr. Samuel Wong

PUBLICATIONS & SUBMITTED WORKS

Yip, B., Bingham, G.*, **Kempfert, K.***, Fabish, J., Kling, T., Chen, C., & Wang, Y. Preliminary Studies on a Large Face Database. *2018 IEEE International Conference on Big Data*. ***These authors contributed equally**

Kempfert, K., Wang, Y., Chen, C., & Wong, S. (2018). Comparison Studies on Nonlinear Dimension Reduction Methods with Kernel Variations: Visualization, Optimization and Classification. Accepted to *Intelligent Data Analysis* to appear in the 2020 issue. Preprint available at <https://arxiv.org/abs/1910.02114>

Kempfert, K. & Wong, S. (2018). Where Does Haydn End and Mozart Begin? Composer Classification of String Quartets. Submitted to *Journal of New Music Research*. Preprint available at <https://arxiv.org/abs/1809.05075>

COMPETITIVE ACADEMIC AWARDS, SCHOLARSHIPS, & FELLOWSHIPS

Chancellor's Fellowship

Aug. 2019 - May 2024

Funds PhD study at the University of California Berkeley for "exceptional applicants who also advance the Regents' goals for diversification of the academy"

Outstanding Four-Year Scholar Award

May 2019

Awarded to less than 12 undergraduate students out of thousands graduating in May 2019 at the University of Florida

Best Oral Presentation

Oct. 2017

Awarded for presentation *Nonlinear dimension reduction with kernel functions for computer vision tasks* at the 13th Annual Regional Mathematics and Statistics Conference at the University of North Carolina Greensboro

Florida Academic Scholarship

Aug. 2015 - May 2019

Awarded to university students in Florida, based on test scores, grades, and community service from high school

RESEARCH PROJECTS

Nonlinear Dimension Reduction for Gender Classification

May 2017 - Feb. 2019

Advisors: Dr. Cuixian Chen, Dr. Yishi Wang, & Dr. Samuel Wong

University of North Carolina Wilmington & University of Florida

- Participated in the 10-week Statistical Data Mining & Machine Learning NSF-REU at the University of North Carolina Wilmington then continued research project for over two years
- Developed a novel machine learning pipeline for the large face database Morph-II; classified over 55,000 photographs in Morph-II as picturing either a male or a female; and reached over 95% cross-validated accuracy (competitive with benchmark)
- Compared the performance of kernel principal component analysis (KPCA), supervised KPCA, and kernel linear discriminant analysis via simulation studies and results on Morph-II

Predicting Classical Composers with Musical Scores

Aug. 2017 - May 2019

Advisor: Dr. Samuel Wong

University of Florida

- Classified the composer of Haydn and Mozart string quartets based on musical scores and set benchmark results that exceed 85% leave-one-out classification accuracy
- Developed novel, musically sophisticated features that can be calculated from musical scores and applied to other music classification tasks
- Generated insights of interest to musicologists and historians through statistical interpretation of results (via feature selection and estimated coefficients in the final model)

Forecasting Dengue Fever in Brazil with Diverse Data Streams

June 2018 - Aug. 2019

Advisors: Dr. Carrie Manore, Dr. Geoffrey Fairchild, Dr. Nidhi Parikh, & Dr. David Osthus

Los Alamos National Laboratory

- Began the project by participating in the 10-week Parallel Computing Summer School and returned the next summer in the Information Systems & Modeling (A-1) research group
- Forecasted dengue fever with high accuracy and confidence for all 27 states of Brazil using time series variables from diverse data streams (doctors' offices, weather stations, satellites, and Google Health Trends)
- Systematically compared predictive performance among variants of seasonal autoregressive integrated moving average (SARIMA), vector autoregression, seasonal trend decomposition, and ensembles combining these methods; reached Pearson correlation coefficients (between observed and fitted values) of up to 96.44% for 2-week-ahead forecasting

TEACHING

Teaching Assistant for Introduction to Statistics (STA 2023)

Aug. 2016 - May 2019

Professors: Maria Ripol, Megan Mocko, & Stephanie Stine

University of Florida

- Led two to three lab sections of 40 to 126 students every week and introduced the main topics of statistics; engaged students in relevant activities, such as simulation of sampling methods, data collection, and use of statistical software Minitab; assigned worksheets for students to complete in lab; graded and recorded labs
- Held office hours every week, providing walk-in tutoring to students; proofread and proctored three exams each semester

ORAL PRESENTATIONS

Heterogenous Data Fusion of Time Series to Nowcast Dengue at the State Level of Brazil. Presented at the Information Systems & Modeling (A-1) research group at Los Alamos National Laboratory, Los Alamos, New Mexico, 2019

Parallel Forecasting of Dengue Fever in Brazil. Presented at the Information Science & Technology Institute's Summer Schools Presentation Day at Los Alamos National Laboratory, Los Alamos, New Mexico, 2018

Kernel-Based Nonlinear Dimension Reduction for Automatic Gender Classification. Presented a contributed paper to Methods for Analysis of High-Dimensional Data session at Joint Statistical Meetings (JSM), Vancouver, British Columbia, 2018

Forecasting Dengue in Brazil with Time Series Modeling in Parallel. Presented at the Theoretical-Division Lightning Talk Series at Los Alamos National Laboratory, Los Alamos, New Mexico, 2018

Nonlinear dimension reduction with kernel functions for computer vision tasks. Won best oral presentation at the 13th Annual Regional Mathematics and Statistics Conference at the University of North Carolina Greensboro, Greensboro, North Carolina, 2017

POSTER PRESENTATIONS

Kernel Variants of Component Analysis and Discriminant Analysis for Gender Classification via Faces. Presented poster at University of Florida Research Computing HiPerGator Symposium, Gainesville, Florida, 2018

Forecasting Dengue in Brazil with Time Series Modeling in Parallel. Presented poster at Los Alamos National Laboratory Student Symposium poster session, Los Alamos, New Mexico, 2018

Kernel-Based Nonlinear Dimension Reduction for Face Analysis. Presented poster at the Mathematical Association of America (MAA) Undergraduate Poster Session at the Joint Mathematical Meetings (JMM), San Diego, California, 2018

Nonlinear Dimension Reduction Using Kernel Representations. Presented poster at University of North Carolina Wilmington Summer Research Showcase, Wilmington, North Carolina, 2017

SERVICE

Diversity Committee of Statistics Graduate Student Association	Aug. 2019 - Present
Service Committee of Statistics Graduate Student Association	Aug. 2019 - Present
Berkeley Artificial Intelligence Lab Mentor	Aug. 2019 - Present
Mentor UF	Aug. 2016 - May 2017

PROGRAMMING & SOFTWARE SKILLS

R, Python, Java, GIS, L^AT_EX