



THE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS

Master's Thesis Presentation

Stephen Ling

Department of Statistics
The University of Chicago

“Comparative Analysis of Global and Local Probabilistic Time Series
Forecasting for Contiguous Spatial Demand Regions”

July 3, 2025, at 9:00 AM
Jones 304, 5747 S. Ellis Avenue

Abstract

This study evaluates three probabilistic forecasting strategies using LightGBM: global pooling, cluster level pooling, and station level modeling across a range of scenarios, from fully homogeneous simulated data to highly heterogeneous real world Divvy bike share demand observed during 2023 to 2024. Clustering was performed using the K-means algorithm applied to principal component analysis transformed covariates, which included time series features, counts of nearby transportation infrastructure, and local demographic characteristics. Forecasting performance was assessed using prediction interval coverage probability (PICP), normalized interval width (PINAW), and the mean squared error (MSE) of the median forecast. The results show that global LightGBM models incorporating station identifiers consistently outperform both cluster level and station level models across most scenarios. These global models effectively leverage the full cross sectional dataset while enabling local adjustments through the station identifier, resulting in superior prediction interval coverage, sharper intervals, and lower forecast errors. In contrast, cluster based models often suffer from residual within group heterogeneity, leading to degraded accuracy. Station level models capture fine grained local dynamics in heterogeneous settings. These findings underscore that global LightGBM models with embedded station identifiers provide a robust, scalable, and computationally efficient framework for transportation demand forecasting. By balancing global structure with local specificity, this approach offers a practical and effective solution for real world mobility applications.