Fusion of Bicycle Data Sources
NaTMEC 2018
Frank Proulx, Toole Design Group
Overview

- Motivation
- Background
- Data Sources
- Methodology
- Results
- Conclusions
Motivation

• We need volume data, network-wide (e.g. to apply our safety performance functions)
• Bike trips are difficult to model
Background: Direct Demand Modeling

(Jones et al. 2010, Wang et al. 2013, Lindsey et al. 2006, Strauss and Miranda-Moreno 2013, ...
Background: Utility-Based Models

- Household & Land-Use Characteristics
- Trip Generation
- Trip Distribution
- Mode Choice
- Network Assignment

- # Trips Beginning and Ending in Each Zone
- # Trips Between Zone-pairs
- # Trips Between Zone-pairs by Each Mode
- Flows/Times on Network Links
Opportunity: Proliferation of Data
Methods: Data Homogenization

- Crowdsourced GPS
- TDM Estimates
- Counts
- Bikeshare

- Time Series Truncation, Averaging, and Binning
- Routing (if represented as O/D flows)
- Map-matching (if GPS traces)
Counts: The Ground Truth

Bicycle Count Locations

- Automated Counters
- Manual Count Intersections
Data Sources: Bikeshare

- Station O-D information
- Timestamps of all trips
- Limited spatial coverage
- $\rho = 0.38$
Data Sources: SFCTA

- County-level demand model
- $\rho = 0.59$
Data Sources: MTC

- Regional travel demand model – focus on longer trips?
- Limited spatial resolution
- \( \rho = 0.11 \)
Data Sources: Strava Metro

- Heavy waterfront and bridge traffic – recreation?
- No routing required
- $\rho = 0.72$
Methods: Geographically Weighted Data Fusion
Bandwidth Tuning

Gaussian Kernel Cross-Validation Results

$h^* = 2400$
Results

PM Peak Volumes

Bicycles/Hour
0
1 - 23
23 - 92
92 - 232
232 - 459
459 - 685
Future Research Needs

• Replication in additional cities
• Incorporation of other emerging data
  – Dockless bikeshare
• Different weighting schemes
• Methods to account for gaps in time, temporal patterns
• Alexei Pozdnukhov for advising
• Strava for providing access to data