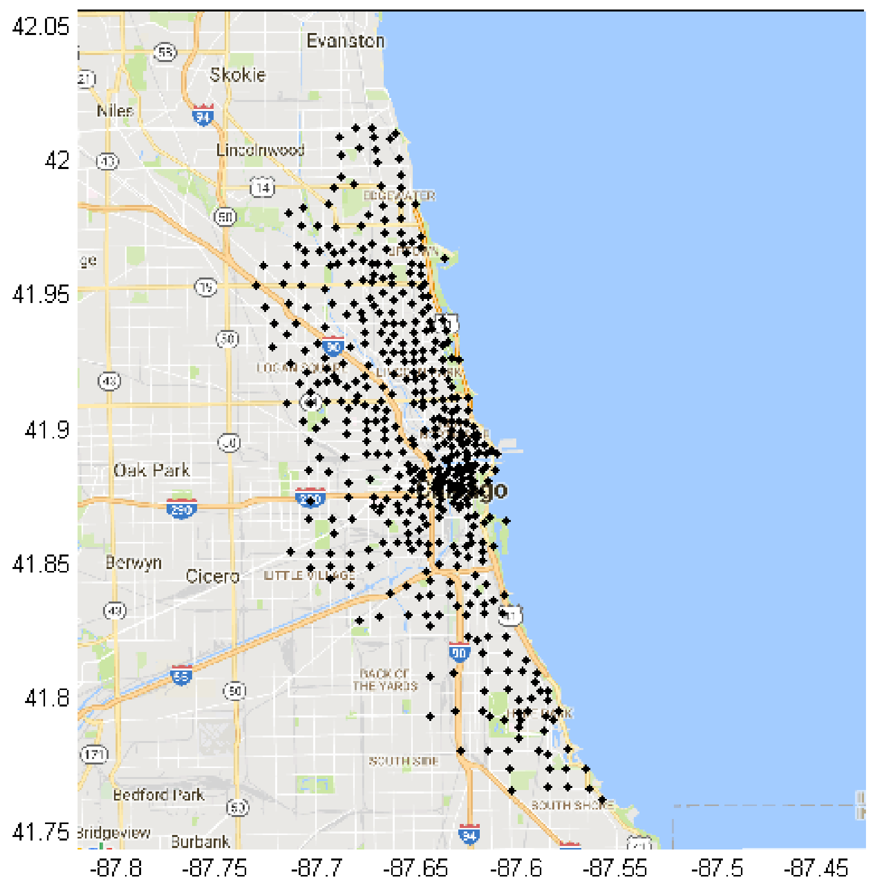
Functional Data Methods for Replicated Point Processes

Daniel Gervini
University of Wisconsin – Milwaukee
Funded by NSF grant DMS-1505780

June 2018

Motivation: Chicago Divvy bike sharing system

458 bike stations active April 1 – November 30, 2016



Chicago Divvy trips 2016

- $3,068,211$ bike trips between April 1 – November 30

- Here we'll look at trip starting times (bike demand)

- This is a spatio-temporal point process on a lattice (since bike station locations are fixed)

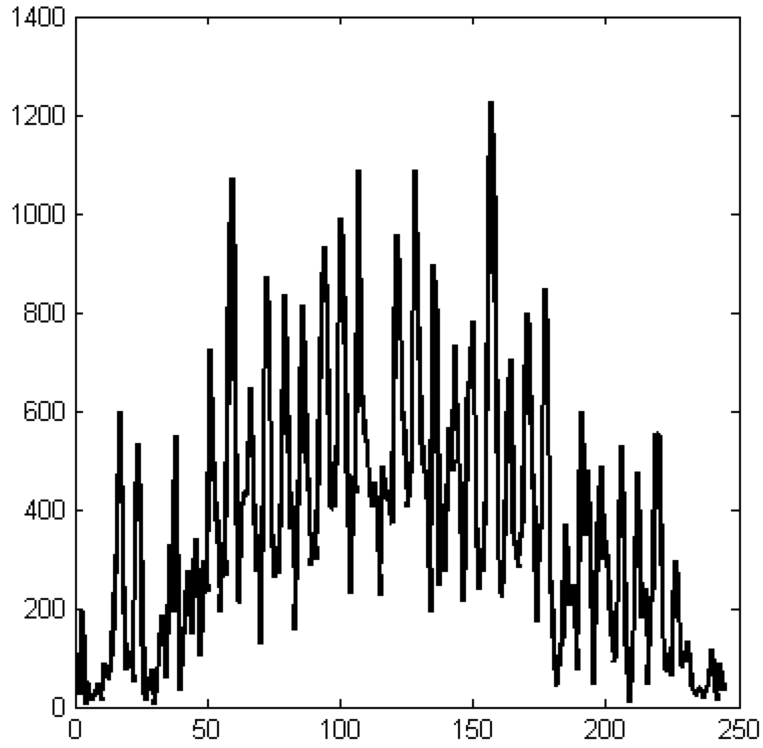
$$\begin{aligned} X_{ij} &= \{\text{trip starting times on day } i \text{ at bike station } j\} \\ i &= 1, \ldots, 244 \\ j &= 1, \ldots, 458 \end{aligned}$$

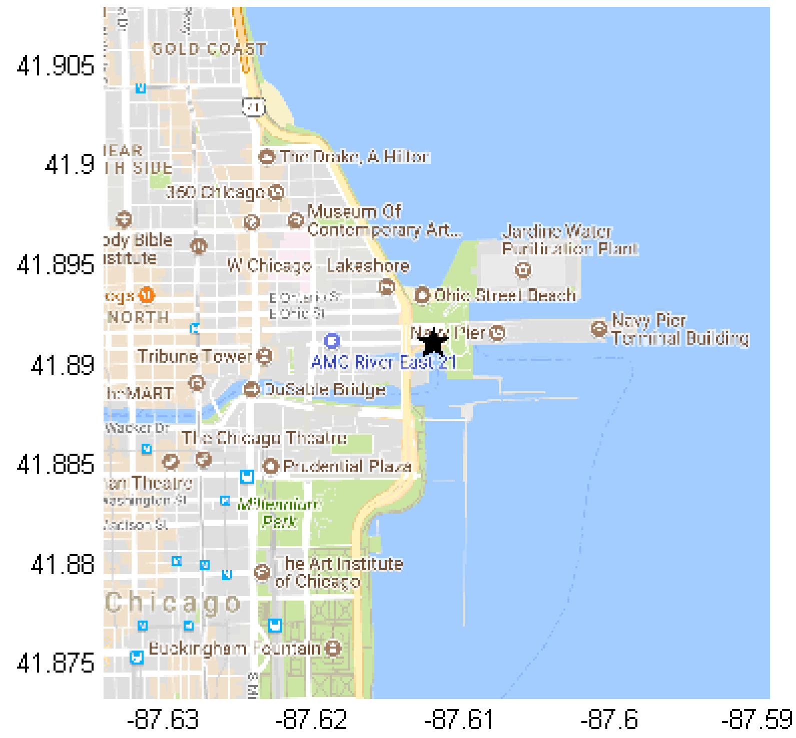
- We can see it as a replicated multivariate temporal process

$$\begin{aligned} \mathbf{X}_i &= (X_{i,1}, \ldots, X_{i,458}) \\ i &= 1, \ldots, 244 \end{aligned}$$

Example

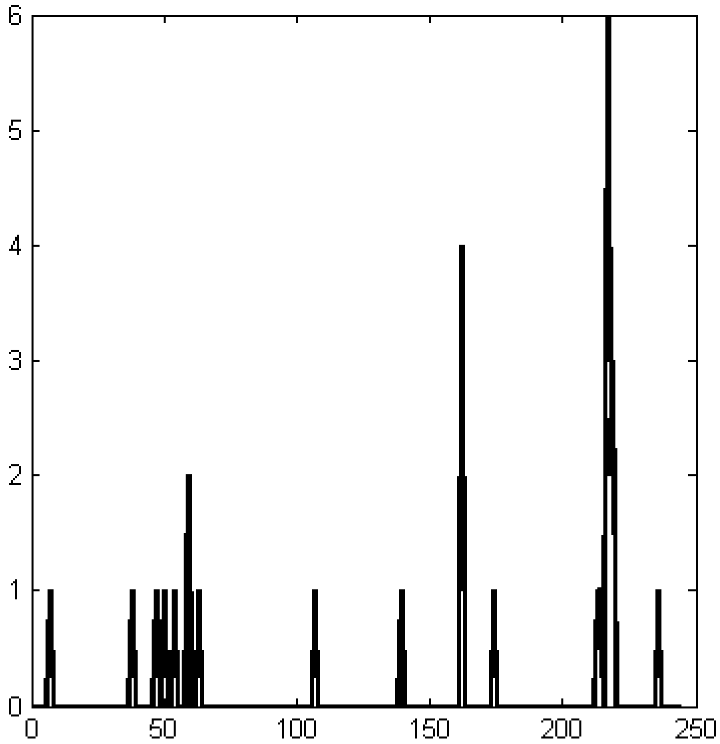
Station with <u>largest</u> total count: $85,314$ (Station id 35, Navy Pier)

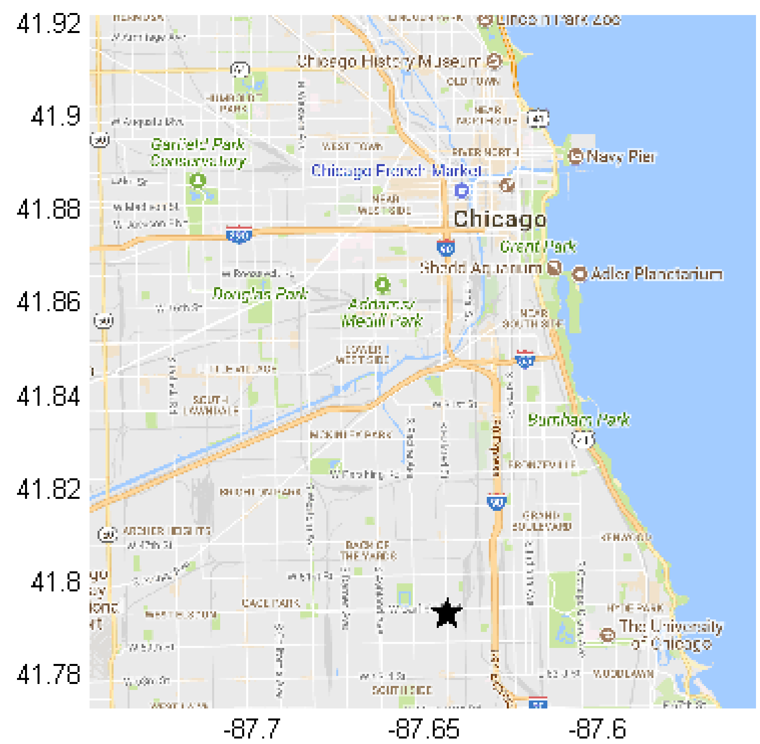




Example

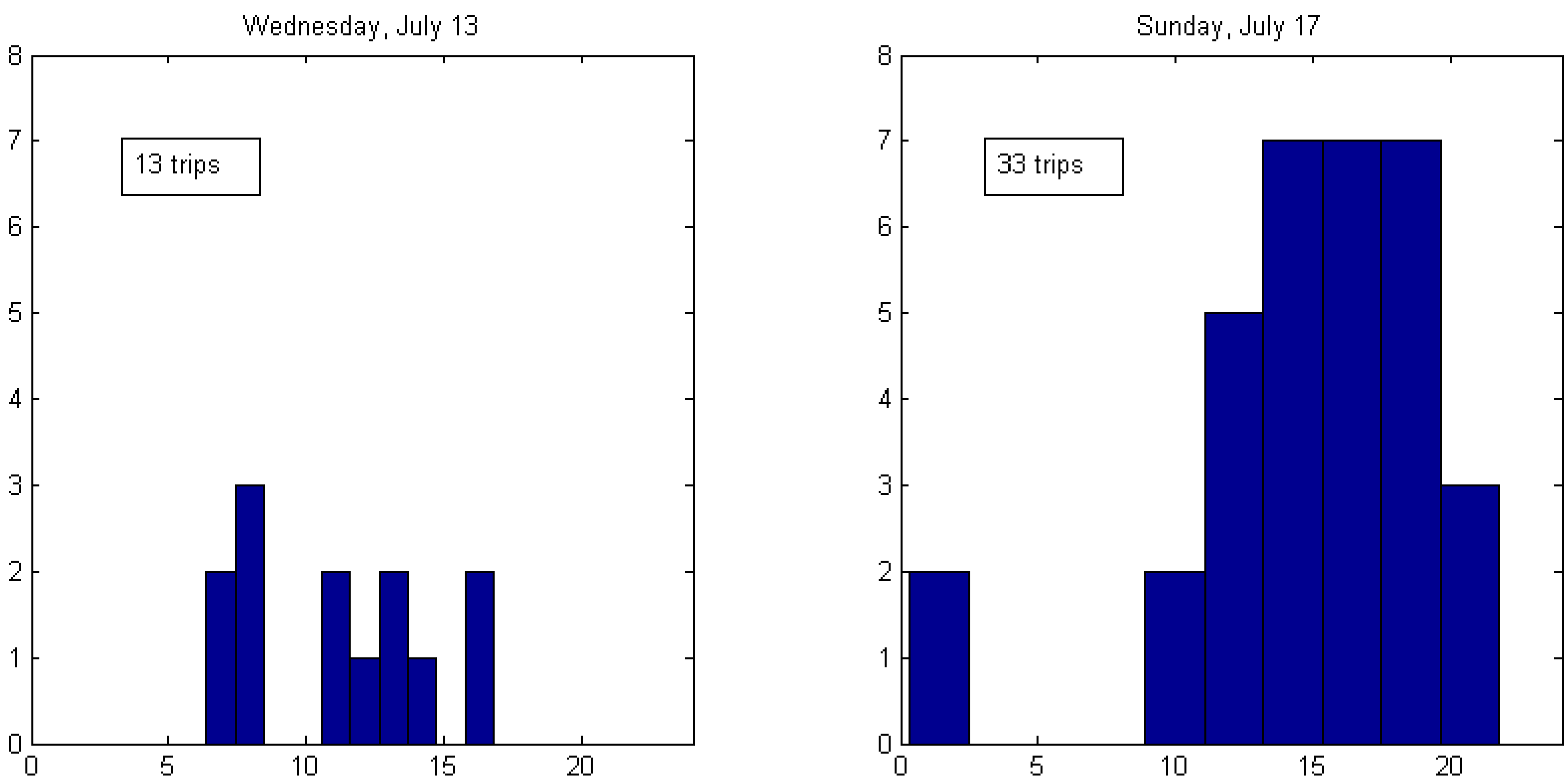
Station with <u>smallest</u> total count: 29 (Station id 386, South Side)





Example

Station with <u>median</u> total count: $4,304$ (Station id 166, Ashland & Wrightwood)



Goals

- Analyze daily variation per station

  - Estimate demand distribution during the day

    (For stations with low counts this can't be done by usual smoothing techniques)

  - Study variation from day to day, for each station

- Analyze spatial variation of bike demand

  - Can bike stations be clustered into groups with similar patterns of bike demand?

  - Do these clusters correspond to natural geographical or demographical neighborhoods?

A primer on Point Processes

- A temporal point process X is a <u>random countable set</u> in $(0, +\infty)$

- X is <u>locally finite</u> if $\#(X \cap B) < \infty$ for any bounded B

- In that case we define the <u>count function</u> $N(B) = \#(X \cap B)$, for any bounded B

- X is a **Poisson process** with <u>intensity function</u> $\lambda(t) \geq 0$ if
  - $N(B)$ has a Poisson distribution with rate $\int_B \lambda(t)dt$
  - Given $N(B) = m$, the m points in $X \cap B$ are independent identically distributed with density $\dfrac{\lambda(t)}{\int_B \lambda}$

Link with Functional Data Analysis

- Suppose we have n replications $X_1, \ldots, X_n$

- Each X_i is associated with an intensity function λ_i

- The λ_is are continuous functions, the usual FDA objects

- The λ_is can be studied/modeled by usual FDA techniques

  (Functional principal components, clustering, regression, etc)

- **But** the λ_is cannot be directly observed and cannot be estimated by usual smoothing techniques if $\#X_i$ is small

  ($\#X_i$ is inherently random, can't be made or assumed big)

Multiplicative Component Analysis (MCA)

- The λ_is are nonnegative, locally integrable functions

- Assuming $\lambda_i > 0$, we fit a PC model on their logarithms

$$\log \lambda_i(t) = \mu(t) + \sum_{k=1}^{p} u_{ik}\phi_k(t)$$

with $\{\phi_k\}_{k=1}^{p}$ orthonormal

- This turns into a multiplicative model for the λ_is

$$\lambda_i(t) = \lambda_0(t) \prod_{k=1}^{p} \phi_k^*(t)^{u_{ik}}$$

with

$$\begin{aligned} \lambda_0(t) &= \exp \mu(t) \\ \phi_k^*(t) &= \exp \phi_k(t) \end{aligned}$$

Maximum likelihood estimation

- We use spline models for the functional parameters:

$$\mu, \phi_k \in \mathsf{span}\{\gamma_1, \ldots, \gamma_q\}$$

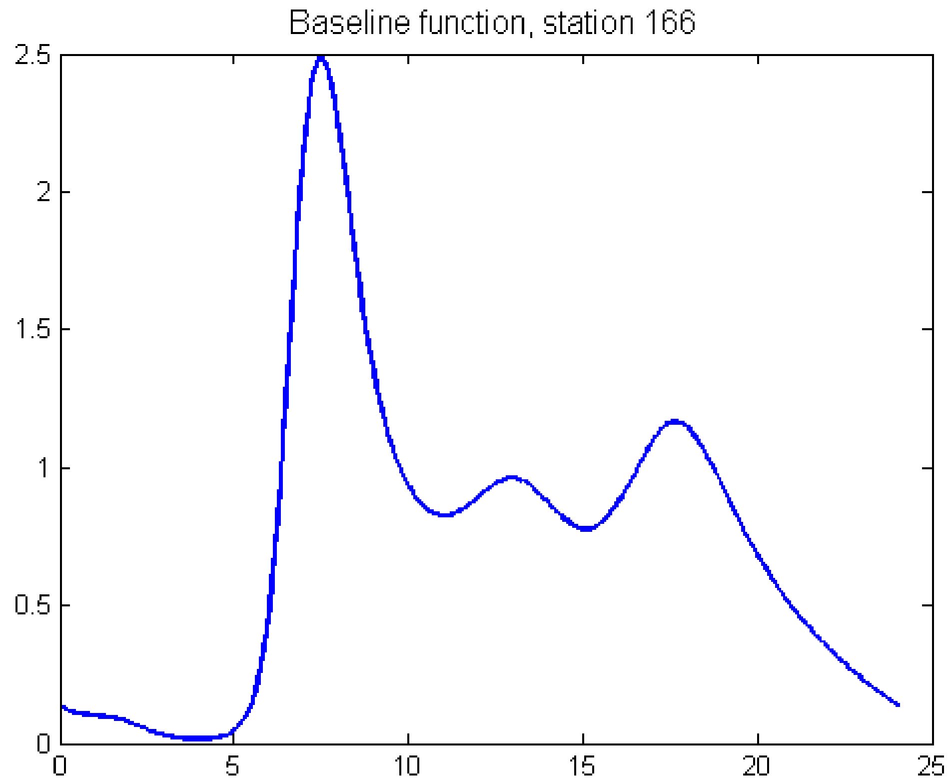
- If $X_i = \{t_1, \ldots, t_{m_i}\}$, we estimate the coefficients of μ and ϕ_ks, and u_{ik}s, by maximizing

$$\begin{aligned} \ell &= \sum_{i=1}^{n} \log f(m_i, t_1, \ldots, t_{m_i}) \\ &= -\sum_{i=1}^{n} \int \lambda_i + \sum_{i=1}^{n} \sum_{l=1}^{m_i} \log \lambda_i(t_l) \end{aligned}$$

- Alternatively, the u_{ik}s can be treated as random effects and integrated out

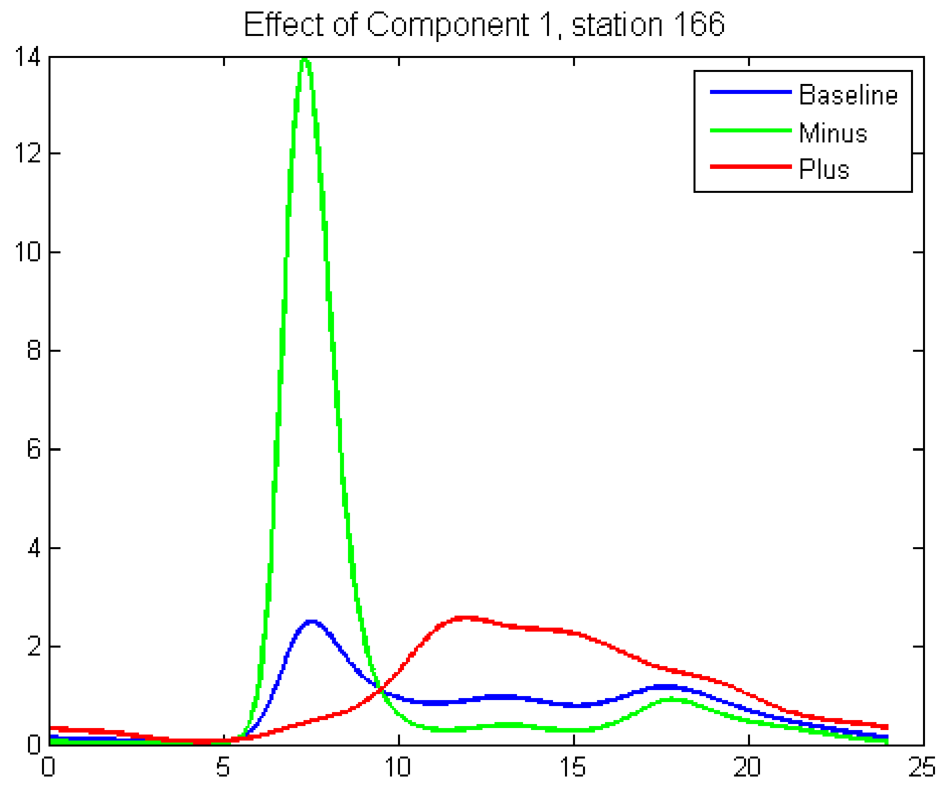
Example: Station 166 (Ashland & Wrightwood)

Baseline intensity function $\lambda_0(t)$



Example: Station 166 (Ashland & Wrightwood)

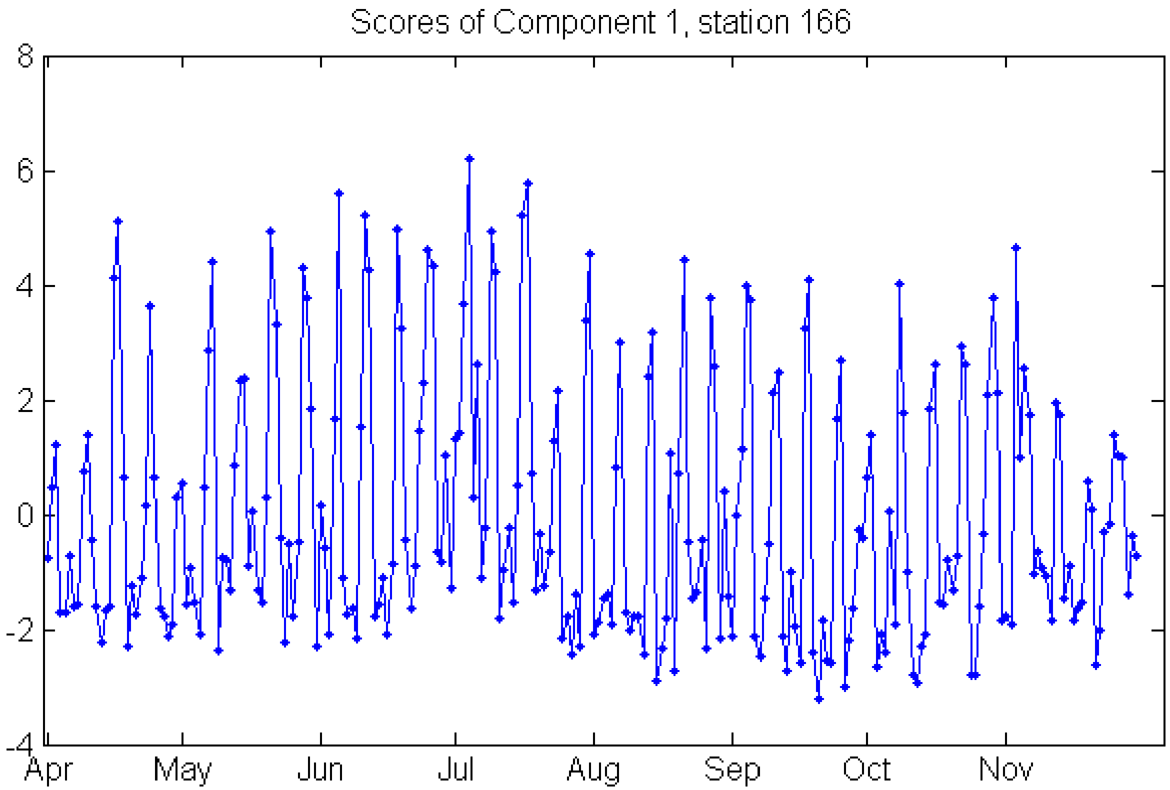
PC 1 is a shape component $\longrightarrow$ morning vs afternoon peak demand



Example: Station 166 (Ashland & Wrightwood)

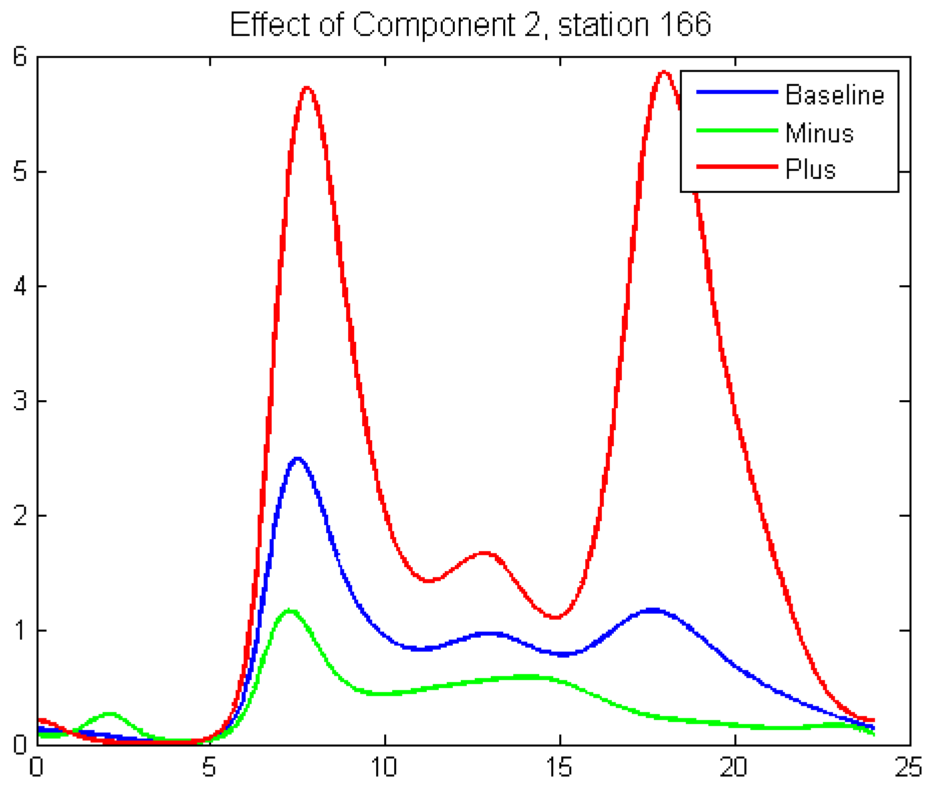
Component scores have no seasonal trend but are weekly periodic

Autocorrelation at lag 7 is .68, peaks occur on Sundays



Example: Station 166 (Ashland & Wrightwood)

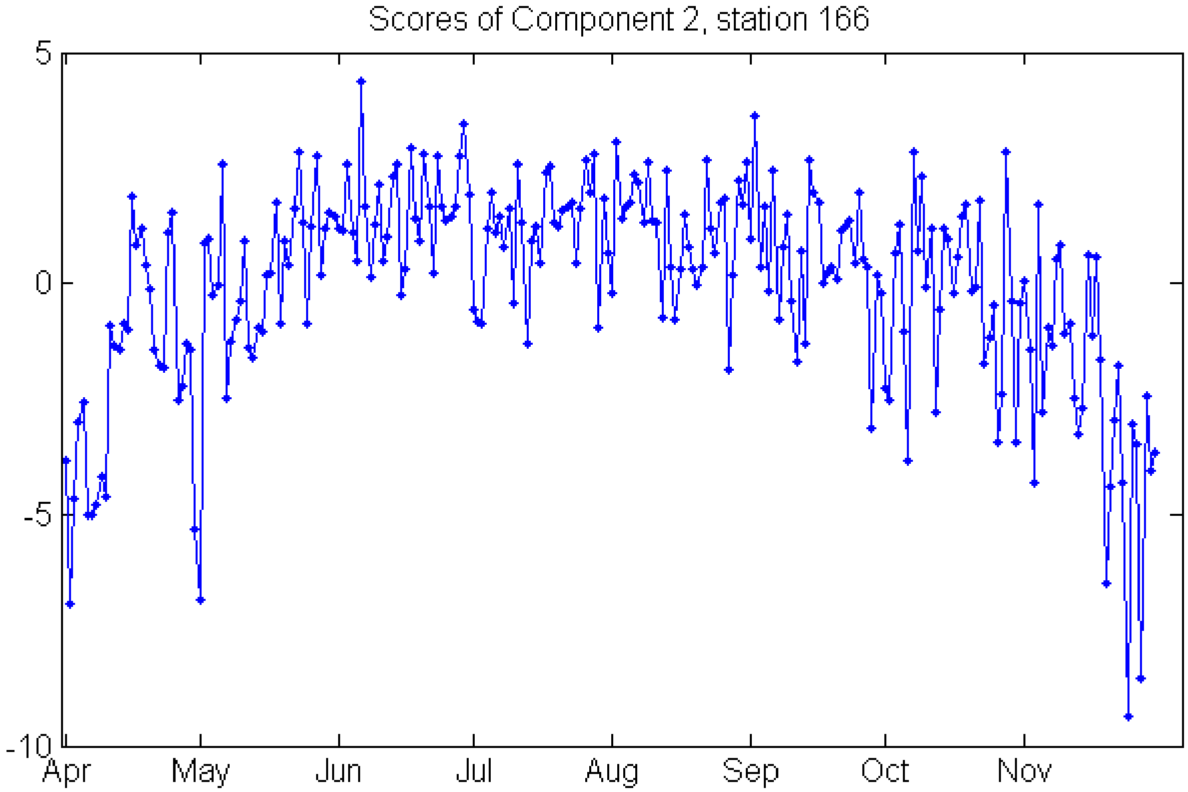
PC 2 is a size component $\longrightarrow$ overall daily count



Example: Station 166 (Ashland & Wrightwood)

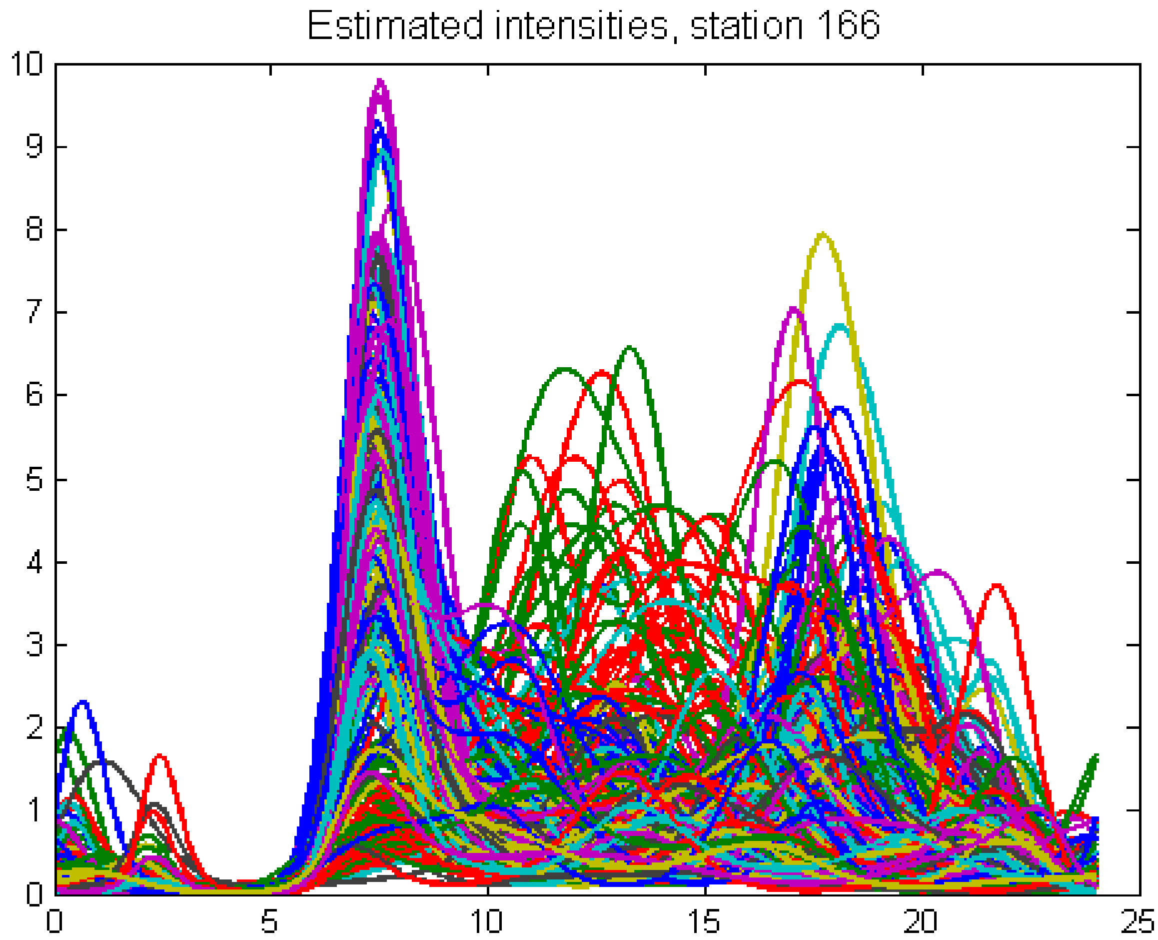
There is a seasonal trend (more bike trips in Summer)

There is some weekly autocorrelation after detrending



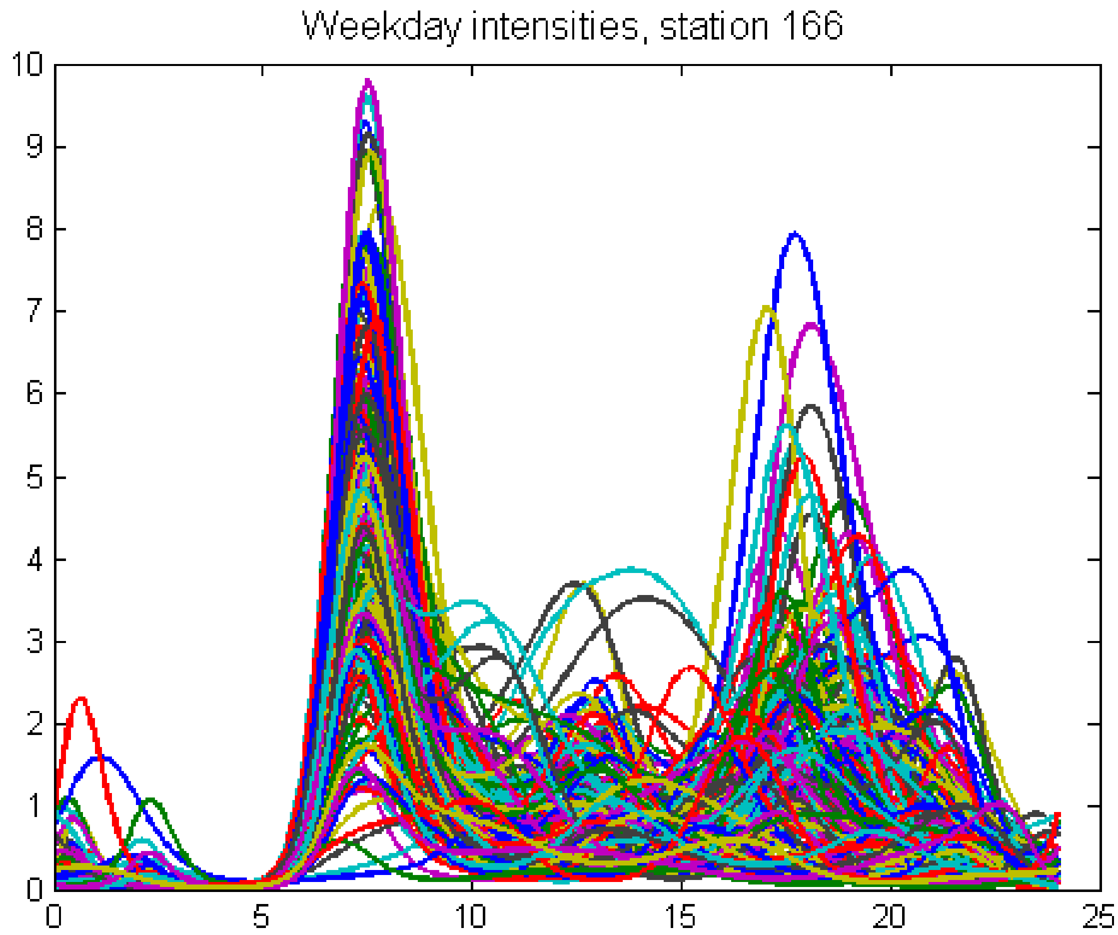
Example: Station 166 (Ashland & Wrightwood)

Estimated daily intensities $\hat{\lambda}_i(t) = \exp\left\{\hat{\mu}(t) + \sum_{k=1}^{p} \hat{u}_{ik}\hat{\phi}_k(t)\right\}$



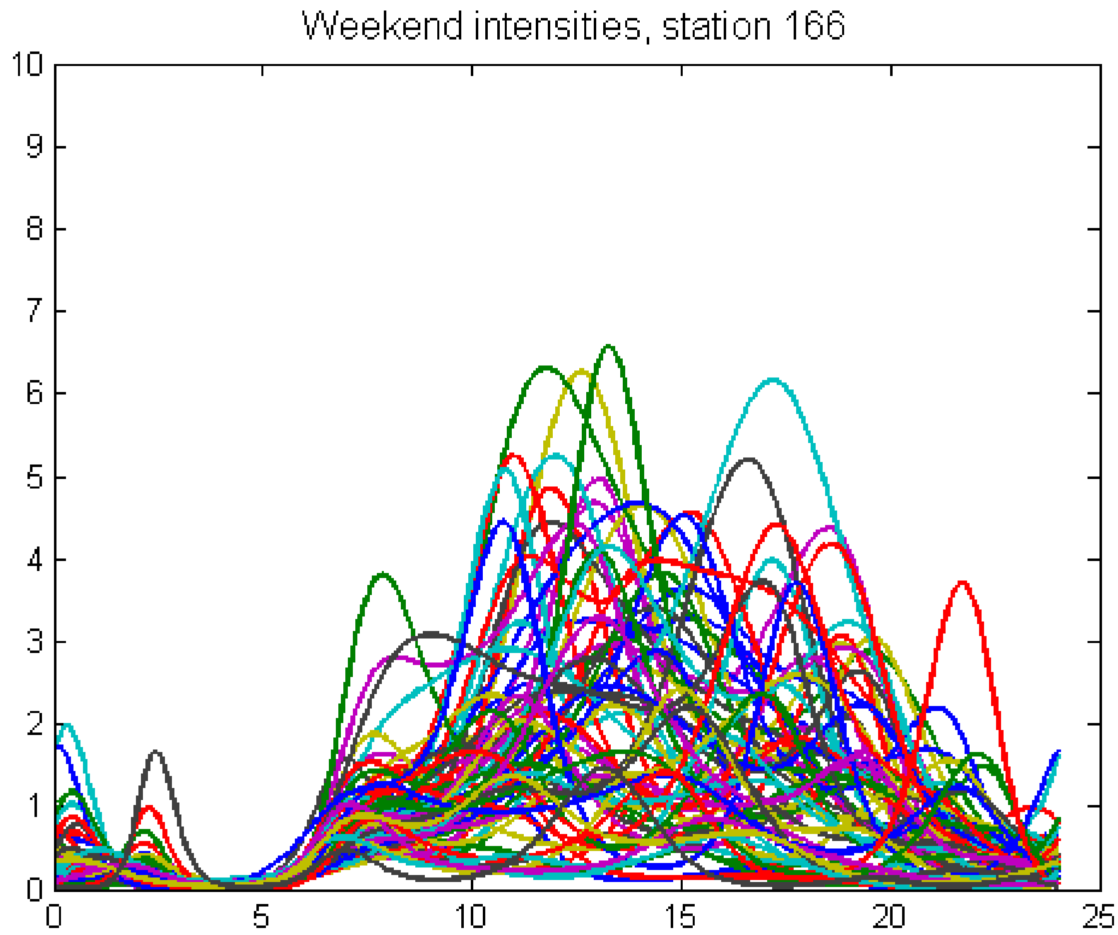
Example: Station 166 (Ashland & Wrightwood)

Weekday $\hat{\lambda}_i$s



Example: Station 166 (Ashland & Wrightwood)

Weekend $\hat{\lambda}_i$s



Covariation models

- Suppose we observe a (functional) variable at each point t_{ij}

$$y_{ij} = g_i(t_{ij}) + \varepsilon_{ij}$$

- Assume $g_1, \ldots, g_n$ are replications of a continuous process G, with PC decomposition

$$g_i(t) = \nu(t) + \sum_{k=1}^{q} v_{ik}\psi_k(t)$$

  with ψ_ks orthogonal

- Association between X and G is studied via correlations between X-scores u_{ik} and G-scores v_{ik}

Estimation

- We treat u_{ik}s and v_{ik}s as latent variables

$$(u_{i1}, \ldots, u_{ip}, v_{i1}, \ldots, v_{iq}) \sim N(\mathbf{0}, \mathbf{\Sigma})$$

- Spline coefficients of μ, ν, ϕ_ks and ψ_ks, and covariance matrix $\mathbf{\Sigma}$, are estimated by (marginal) maximum likelihood

- Inference is based on cross-correlation matrix $\mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1/2}$

- Regression models can be set up:

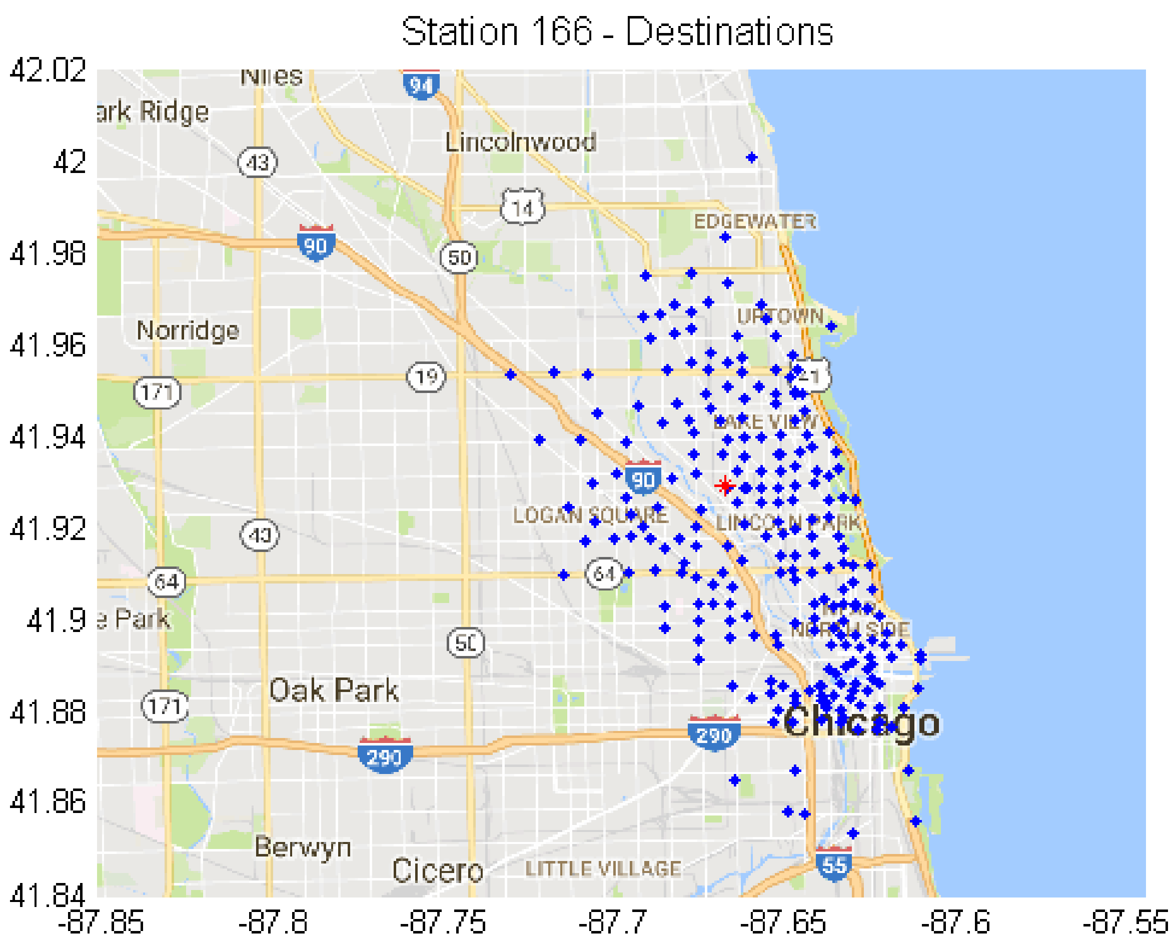
$$\hat{\mathbf{u}} = \mathbf{R}_1\mathbf{v} \quad \text{or} \quad \hat{\mathbf{v}} = \mathbf{R}_2\mathbf{u}$$

  with

$$\mathbf{R}_1 = \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1} \quad \text{or} \quad \mathbf{R}_2 = \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}$$

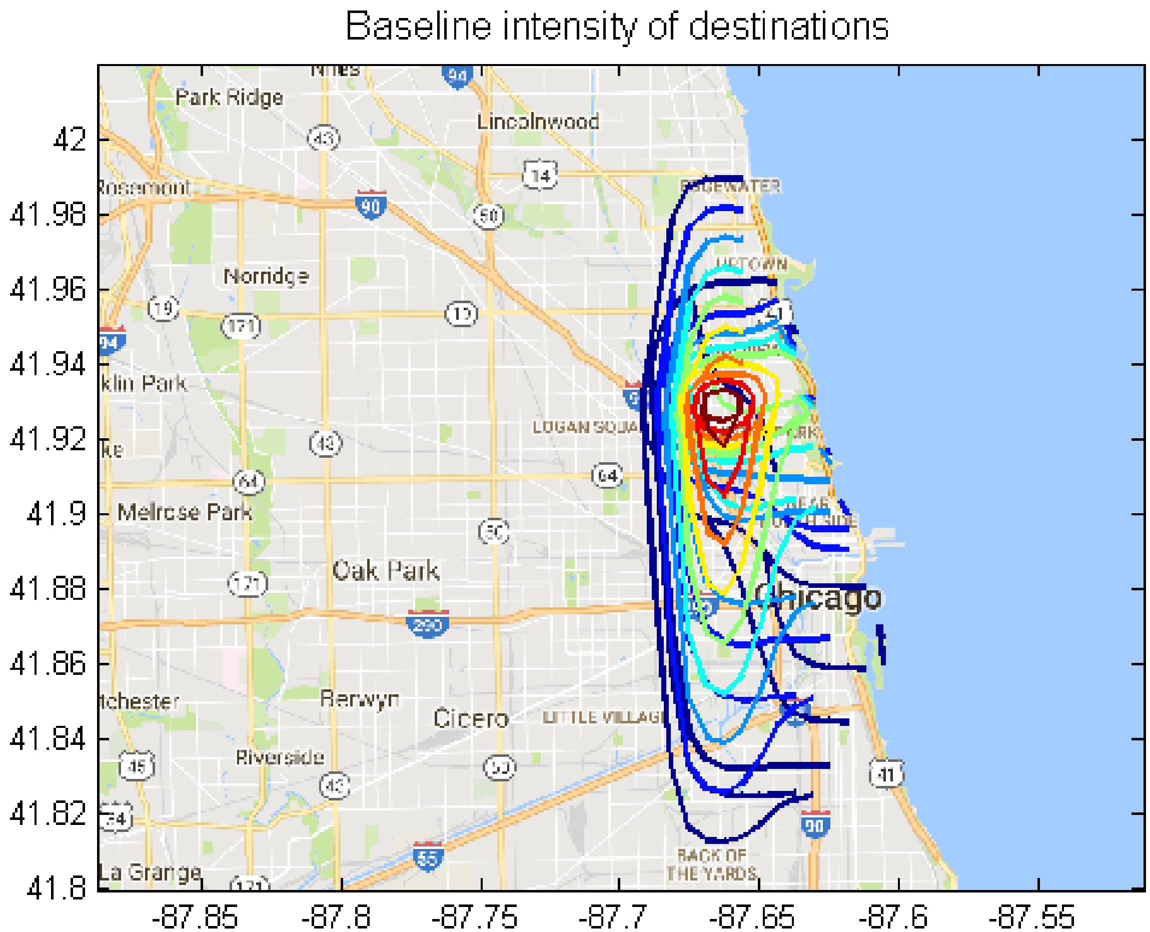
Example: Station 166 destinations

Trip destinations (red dot is Station 166)



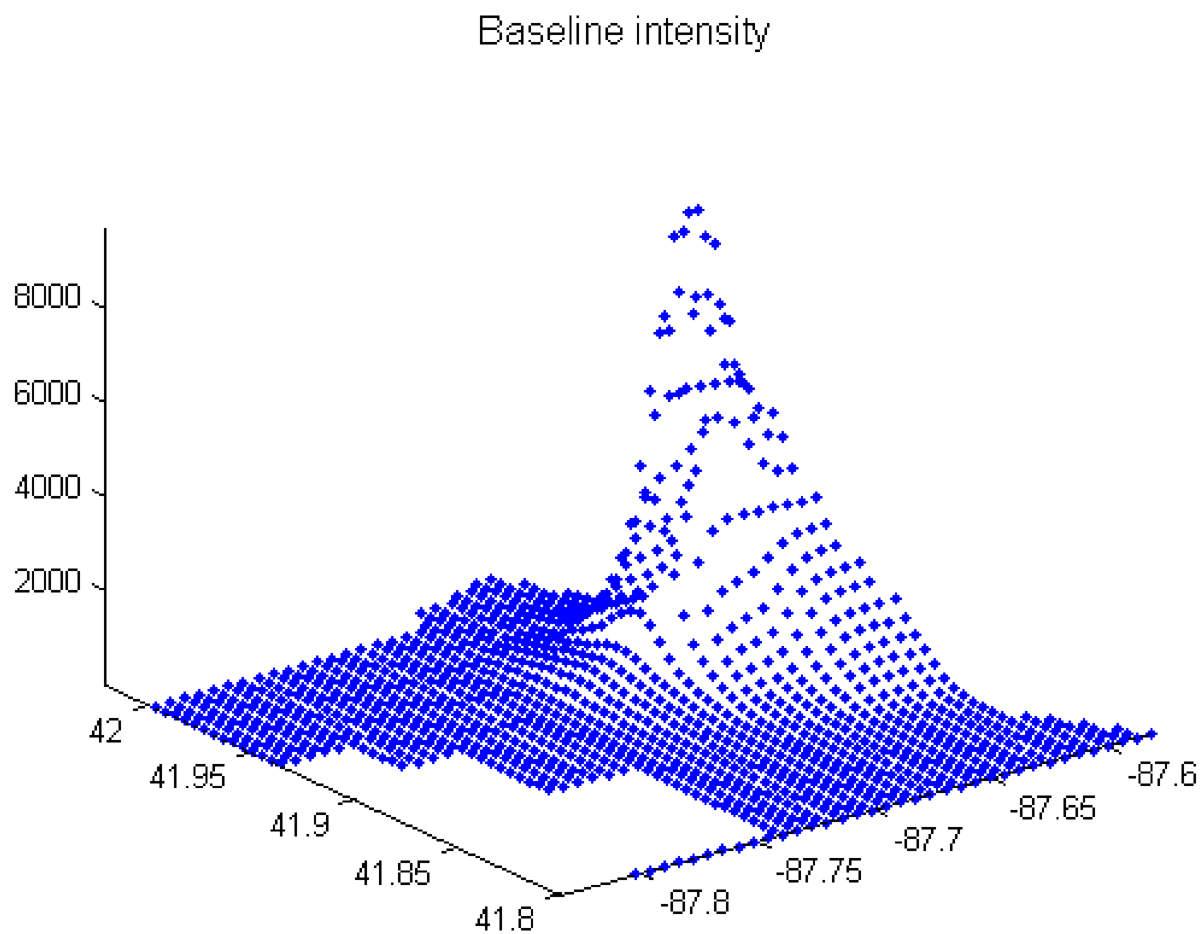
Example: Station 166 destinations

Baseline intensity $\hat{\lambda}_0$



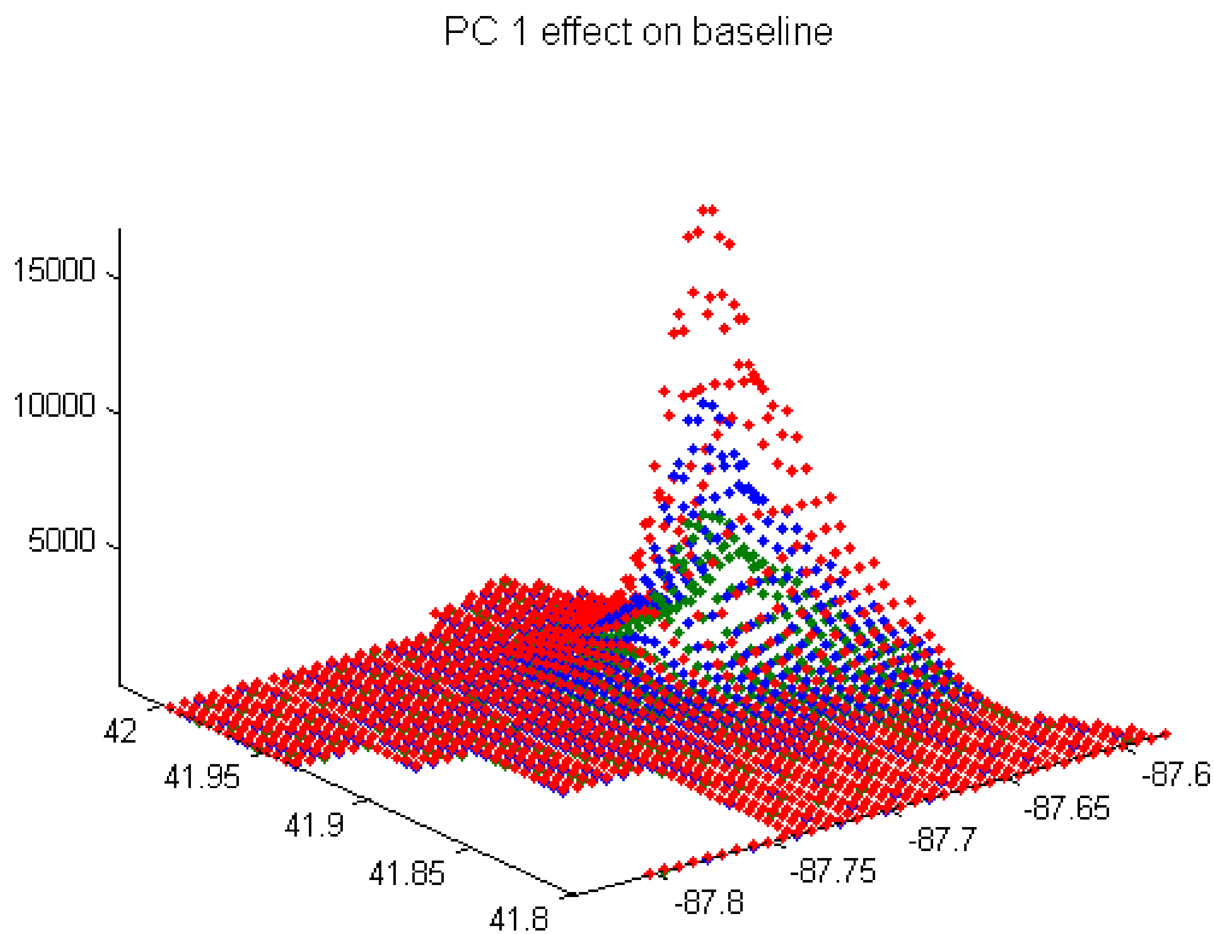
Example: Station 166 destinations

Baseline intensity $\hat{\lambda}_0$



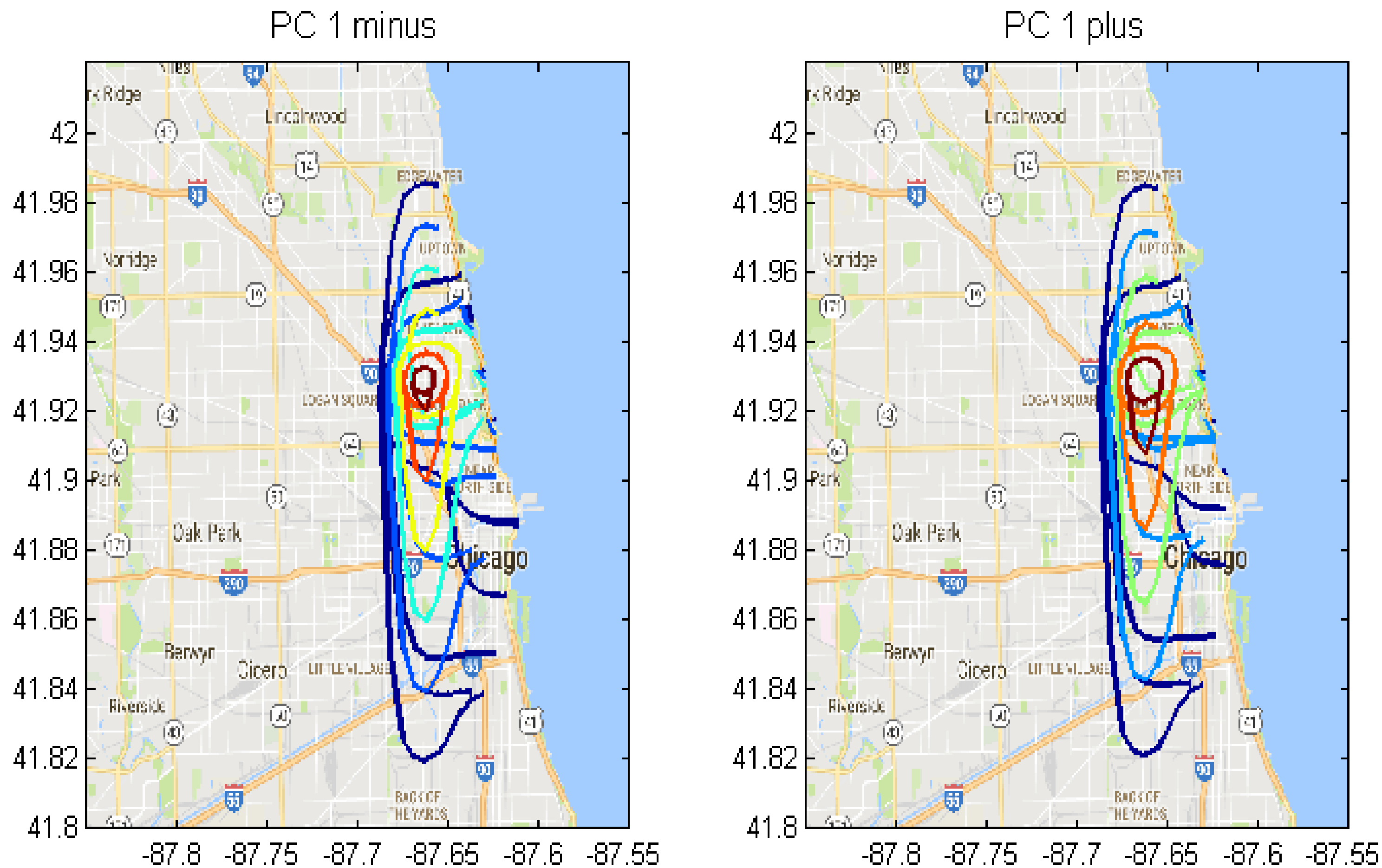
Example: Station 166 destinations

PC 1 $\longrightarrow$ size component (overall count)



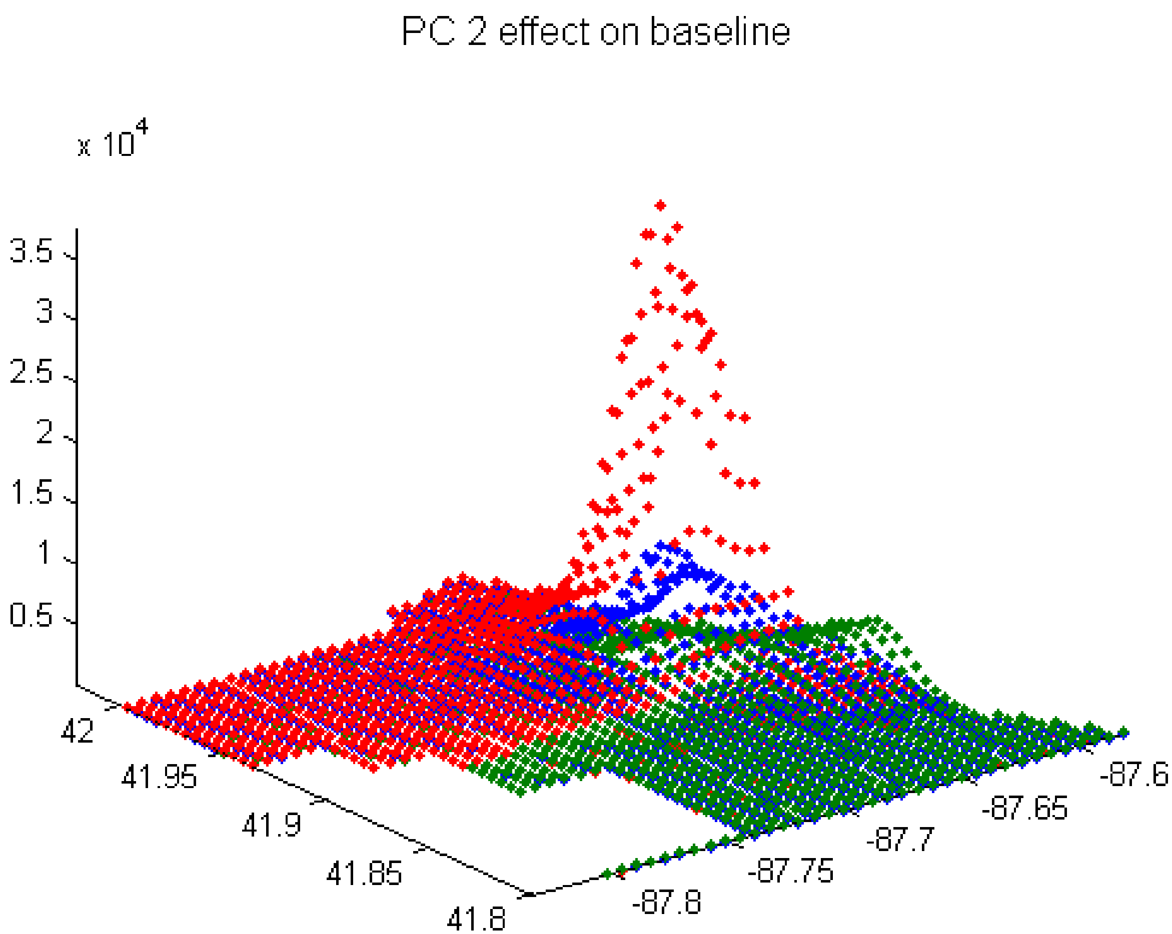
Example: Station 166 destinations

PC 1 $\longrightarrow$ size component (overall count)



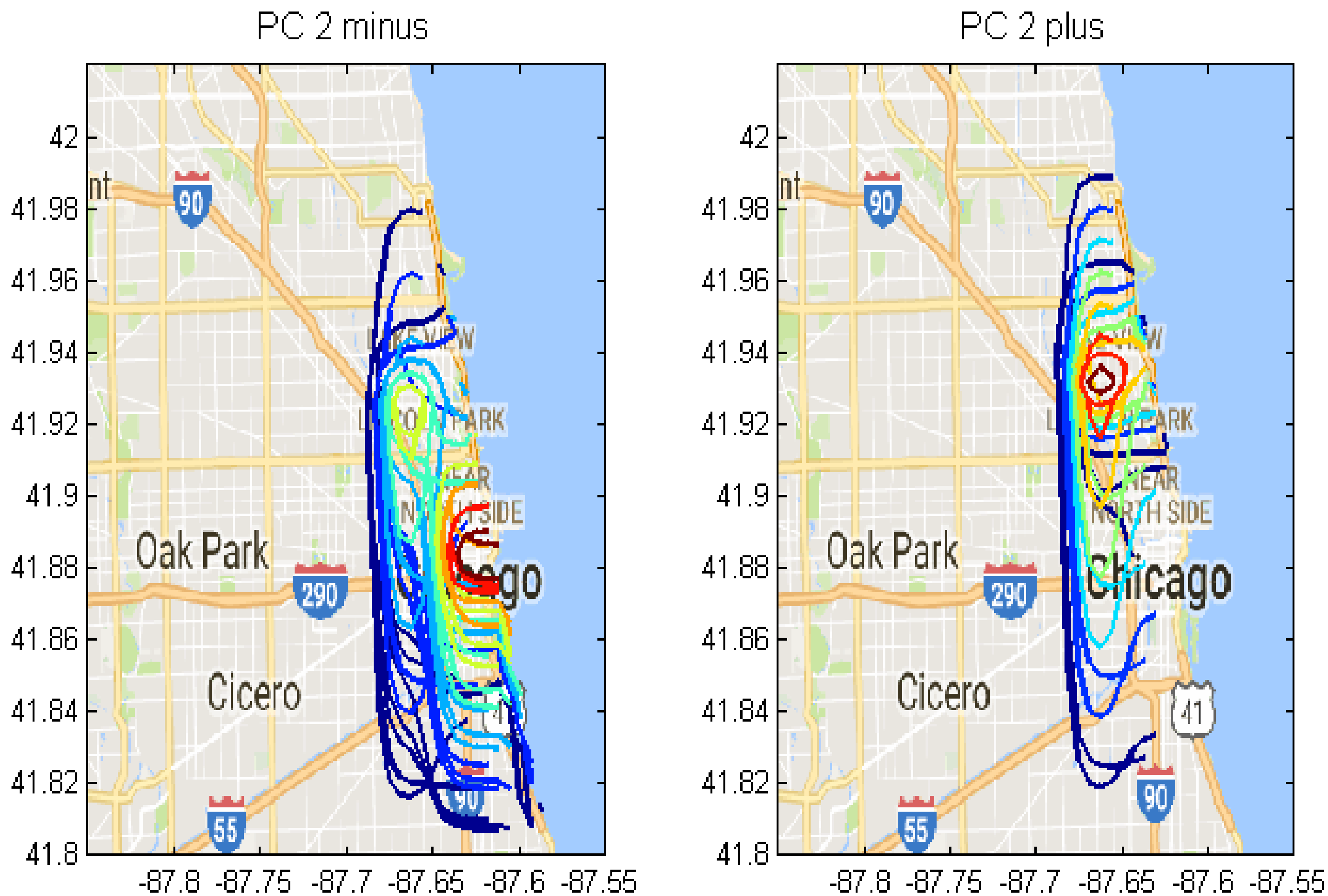
Example: Station 166 destinations

PC 2 $\longrightarrow$ contrast (local vs downtown trips)



Example: Station 166 destinations

PC 2 $\longrightarrow$ contrast (local vs downtown trips)



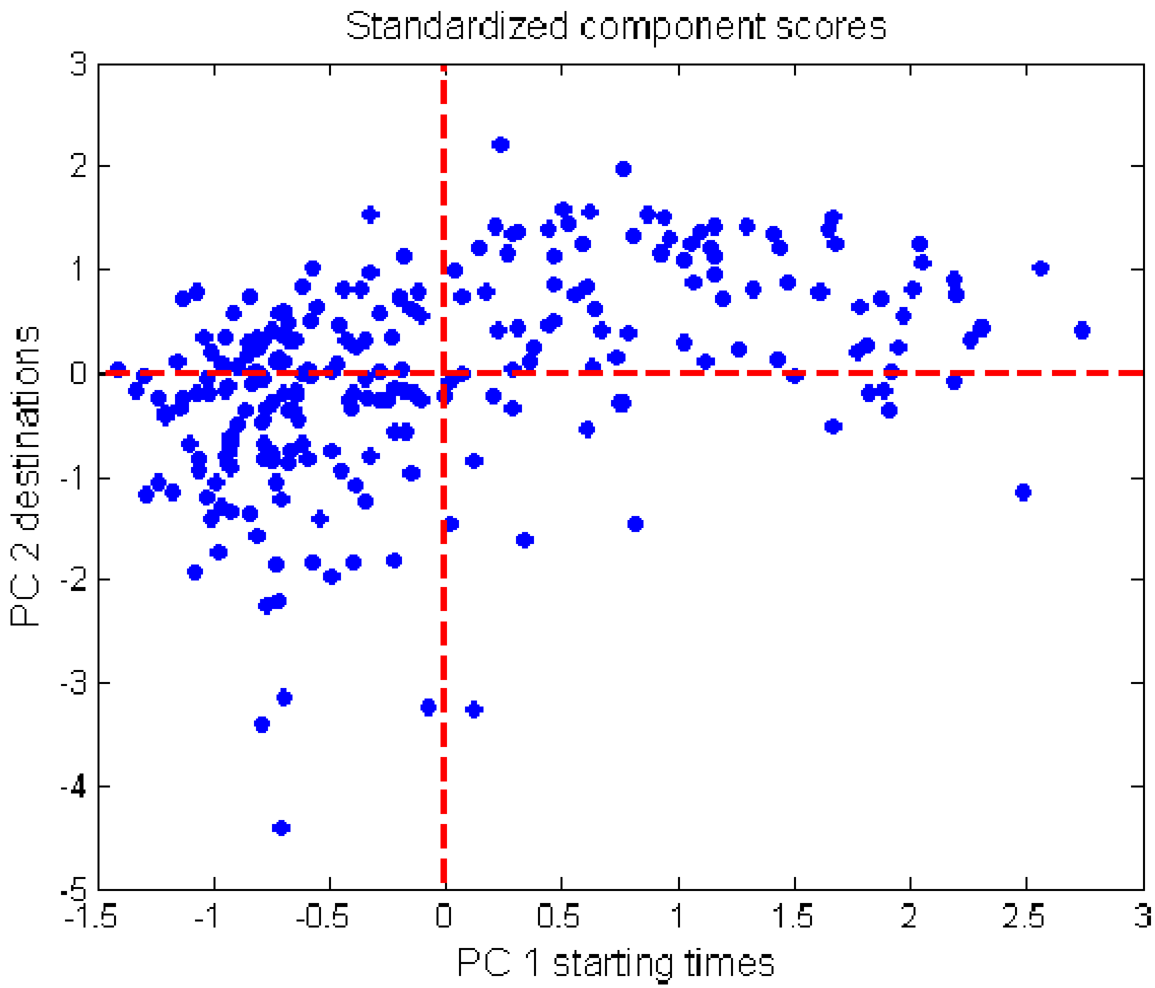
Example: Starting times vs destinations

Cross-correlation matrix of PC scores

Starting time	**Destination**	
	overall size	downtown (–)/local (+)
morning (–)/afternoon (+)	.15	**.43**
overall size	**.91**	.11

Example: Starting times vs destinations

Scatter plot of standardized $\hat{u}_{i1}$s vs $\hat{v}_{i2}$s



Multivariate clustering

- Consider trip starting times for all bike stations, $j = 1, \ldots, 458$

- For two bike stations j_1 and j_2, the functional sample canonical correlation coefficient is

$$\rho_{j_1 j_2} = \max_{\alpha, \beta \in L^2} \operatorname*{corr}_i \left(\left\langle \alpha, \log \lambda_{ij_1} \right\rangle, \left\langle \beta, \log \lambda_{ij_2} \right\rangle \right)$$

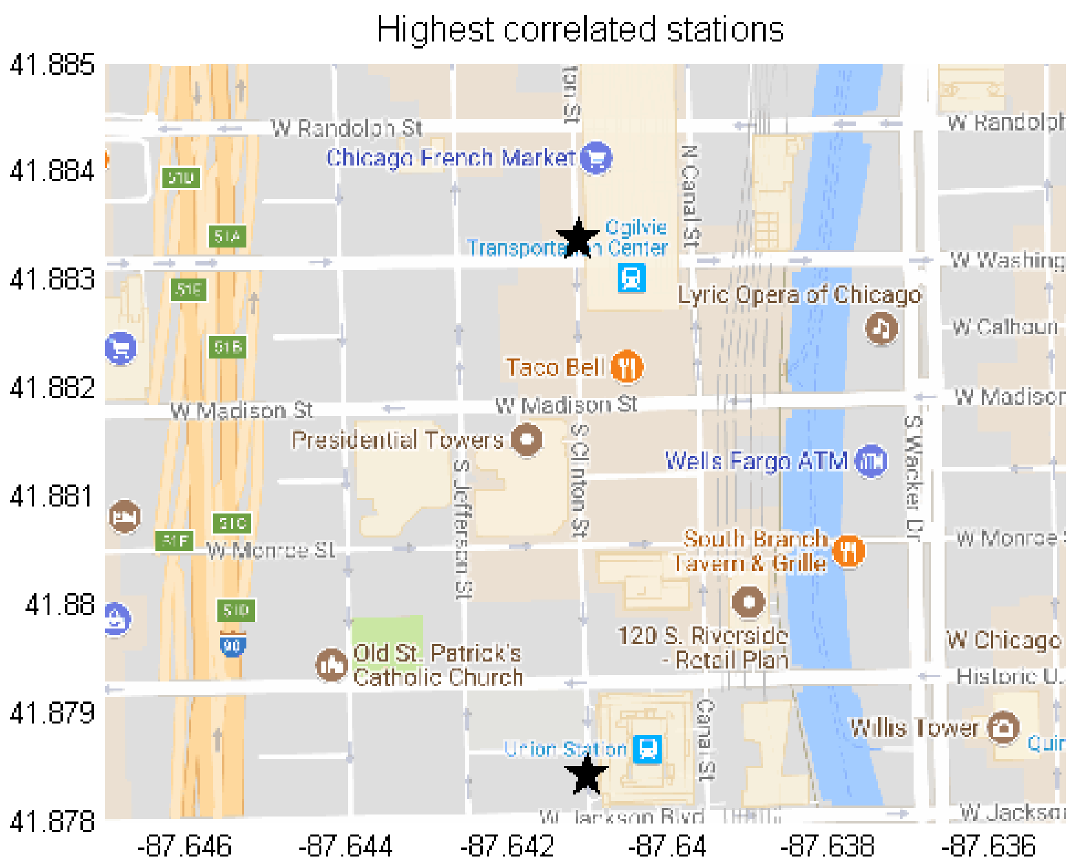
- $\rho_{j_1 j_2}$ comes down to the classical multivariate canonical correlation coefficient of the component scores $\{u_{ikj_1}\}$ and $\{u_{ikj_2}\}$

$$\rho^2_{j_1 j_2} = \max \text{eigen } S^{-1/2}_{j_1 j_1} S_{j_1 j_2} S^{-1}_{j_2 j_2} S_{j_2 j_1} S^{-1/2}_{j_1 j_1}$$

Example: Highest correlated stations

- Largest ρ is .98 for stations 75 (Union) and 91 (Ogilvy)

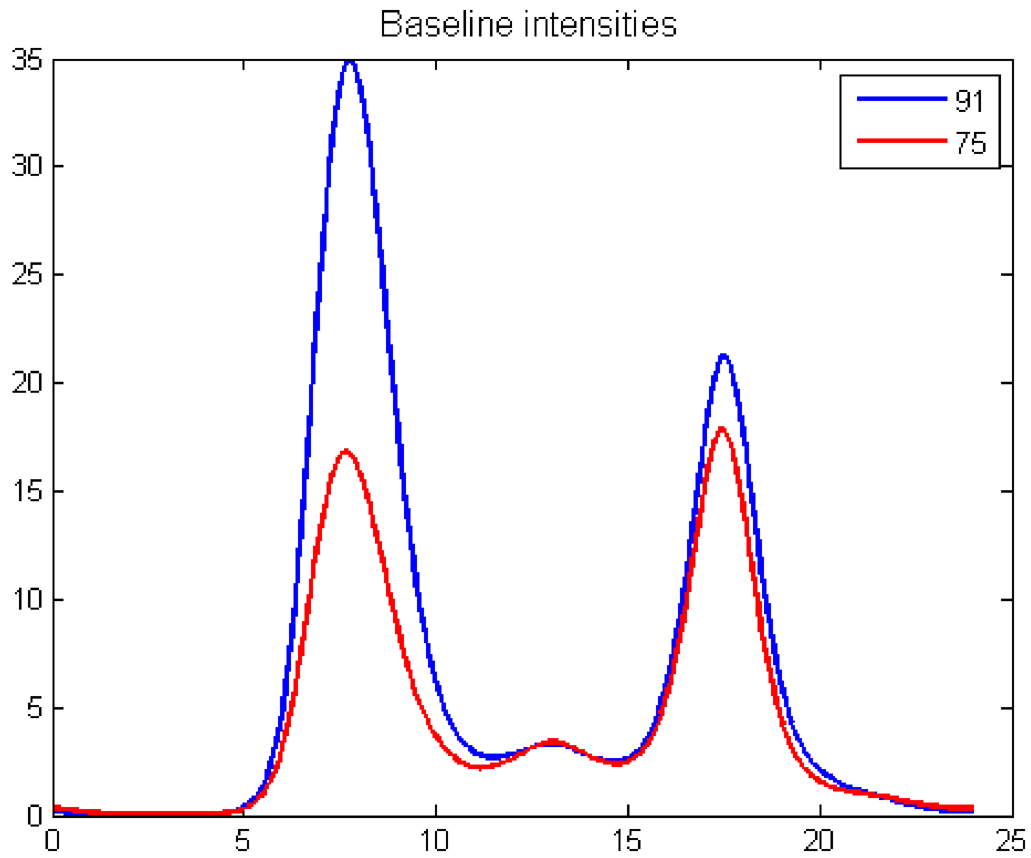
- They are geographically close, but not the closest



Example: Highest correlated stations

- Baseline intensities have similar shapes

- But they are not the closest in L^2 norm

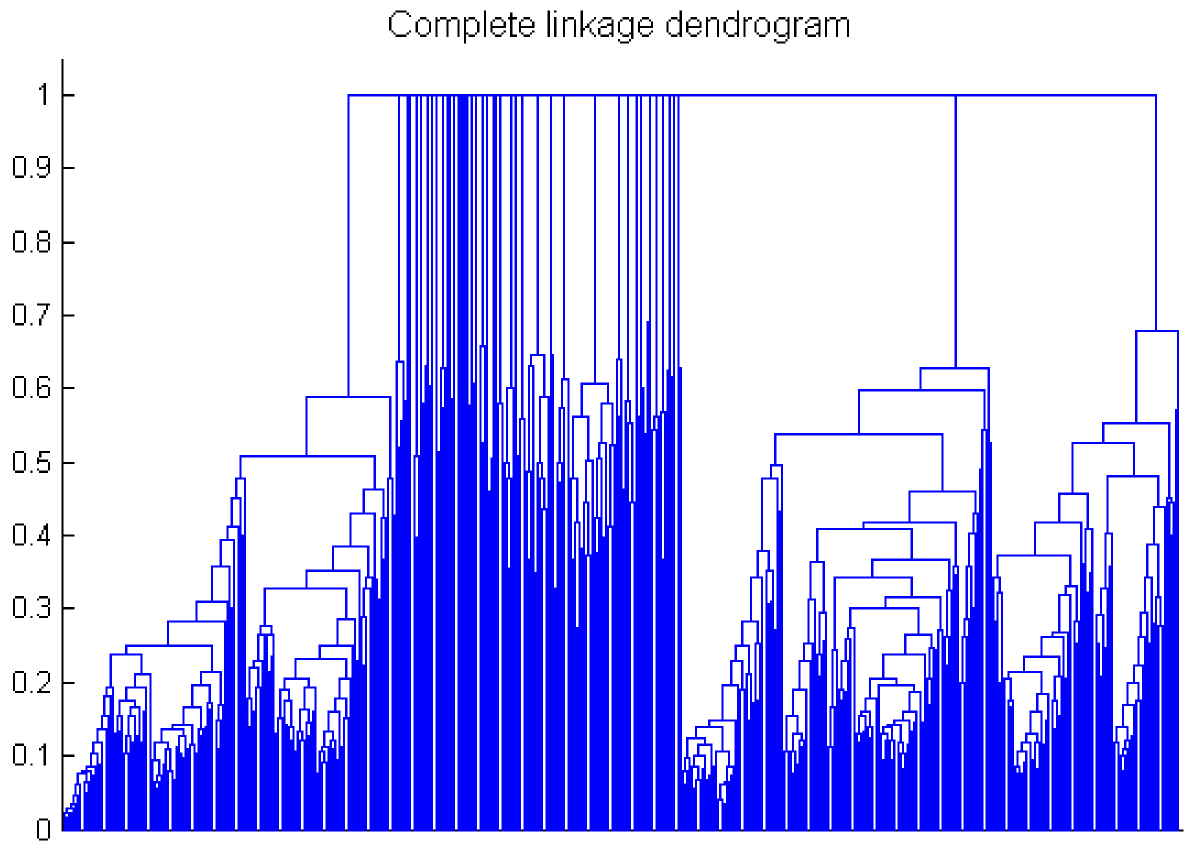


Spatial clustering by correlations

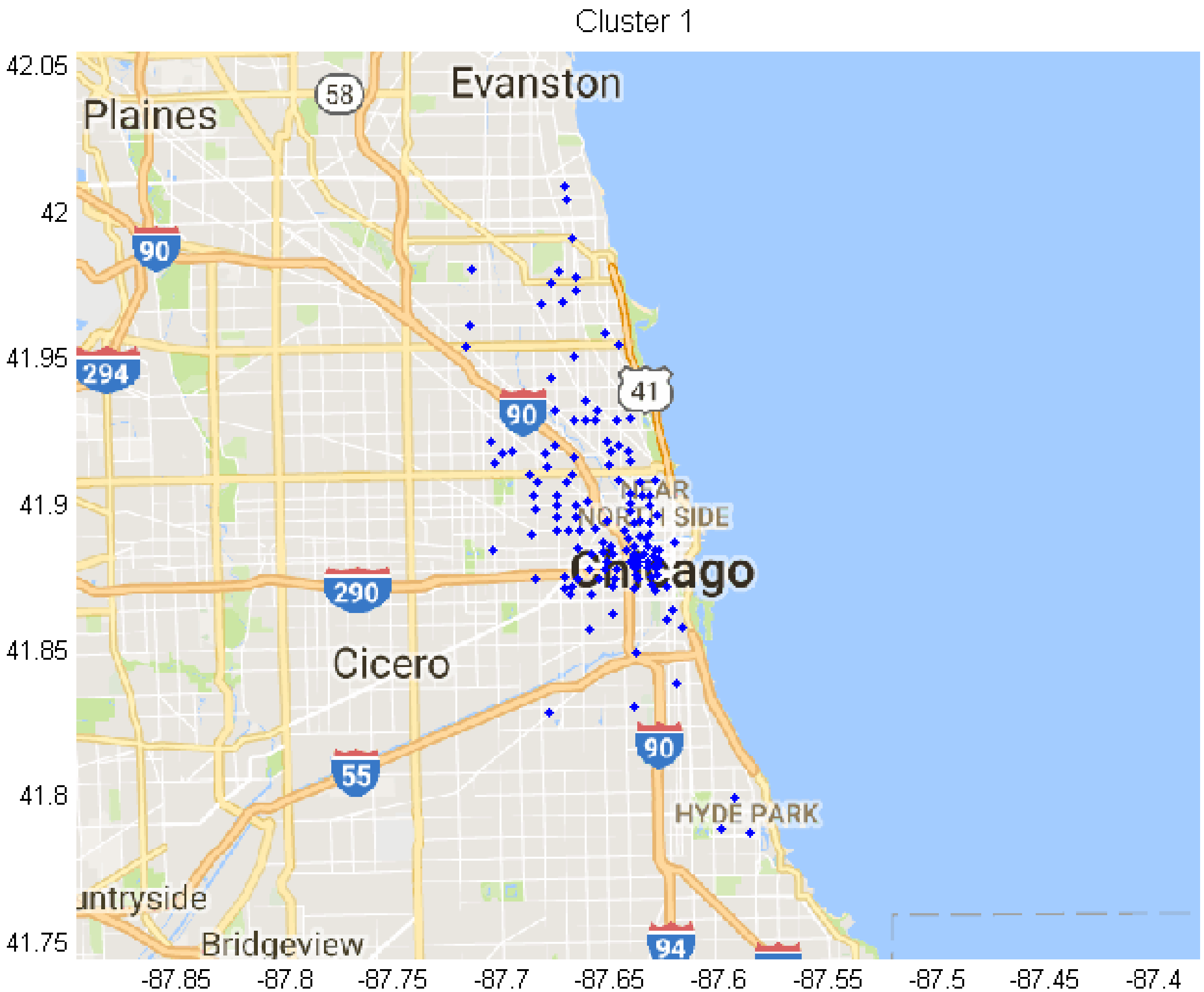
- Filter out non-significant ρ's
  - There are $104,196$ distinct pairs (j_1, j_2)
  - We use Benjamini–Hochberg simultaneous level $\alpha = .01$
- Define distance between bike stations: $d(j_1, j_2) = 1 - \rho_{j_1, j_2}$
- Use agglomerative clustering technique
  - Complete linkage looks best
- Decide how many clusters to retain

Example: Dendrogram for bike stations

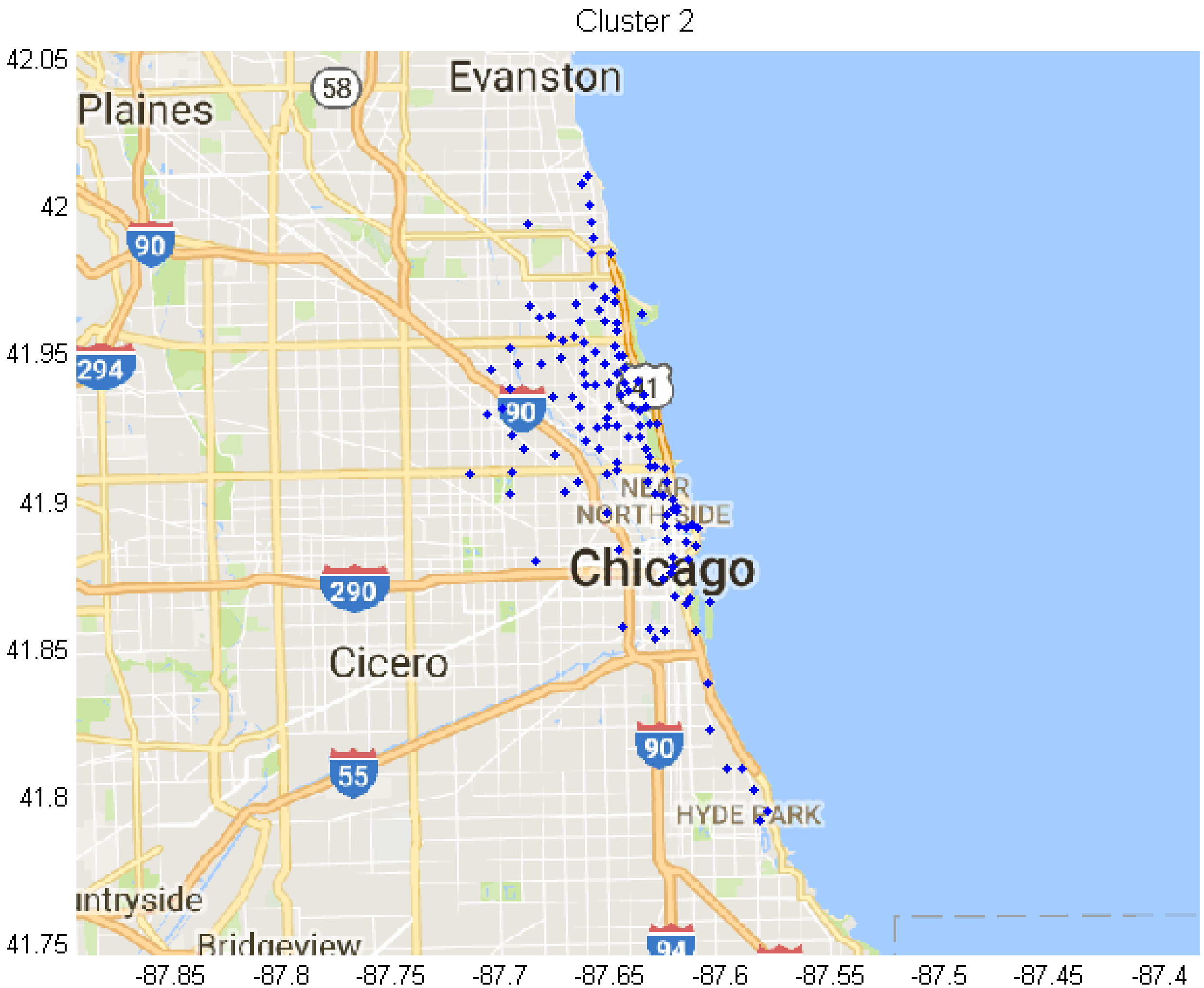
There are three large clusters (136, 127 and 77 stations each)



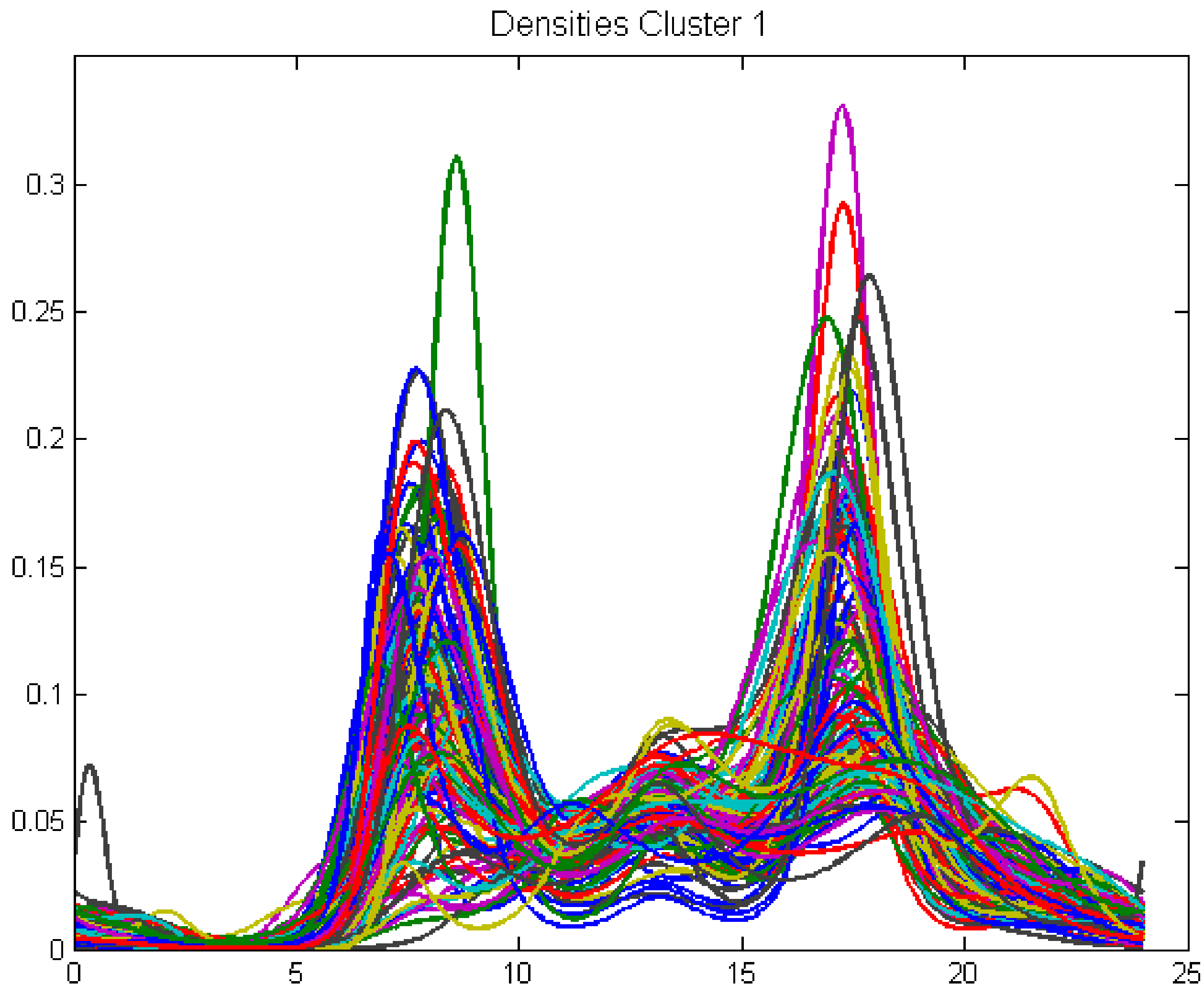
Example: Cluster 1 $\longrightarrow$ "Downtown, Commute"



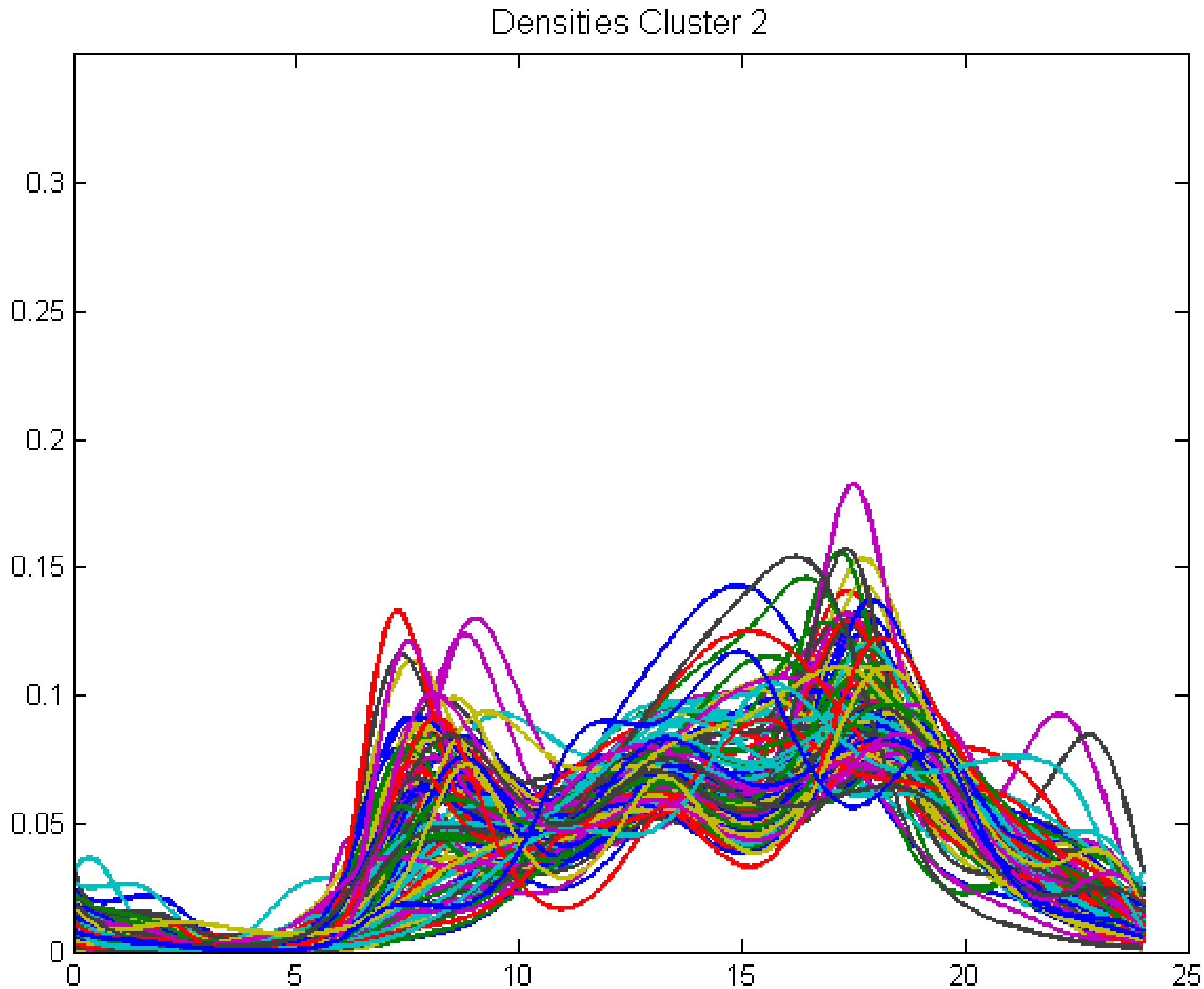
Example: Cluster 2 $\longrightarrow$ "Lakeshore, Leisure"



Example: Densities cluster 1



Example: Densities cluster 2



Thanks!