
Traffic on the I-5, Between Routes 14 and 99

May 6, 2008

Jan de Leeuw and Irina Kukuyeva

Goals of the Project

- n Forecast traffic on the Interstate-5 Freeway Between Routes 14 and 99
- n In our models, we would like to incorporate:
 - q population growth
 - q construction near the freeway

Outline

- n Overview of datasets and previous analyses
 - q *Note:* The talk will omit discussing data management
- n Further research
 - q outlined for each dataset
- n Where to get data

Part A: Dataset Overview

Part 1: AADT

Background

- n Definition: AADT – **A**verage **A**nnual **D**aily **T**raffic
 - n Calculation of AADT (According to AASHTO Guidelines for Traffic Data Programs, The American Association of State Highway and Transportation Officials, 1992, p.52):
 - n *Step 1*: “Averages are calculated for each day of the week ... [and] for each month, yielding seven values for each month”
 - n *Step 2*: "At the end of the year, each of these values is averaged across the twelve months, yielding seven annual values”
 - n *Step 3*: AADT=average of the seven values in Step 2
-

Background

- n 25 interchanges
- n Dataset 1: 1995-2005: no missing data
- n Dataset 2: 1973-2005: 4.8% missing data

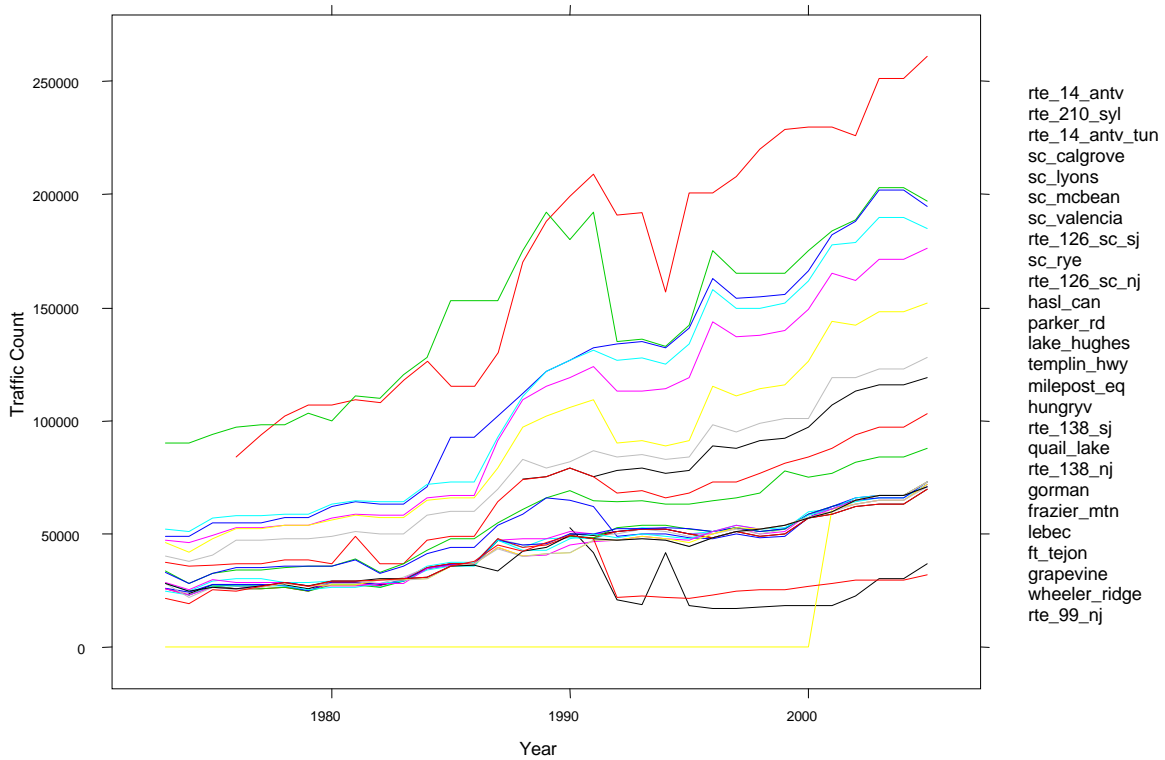


Figure 2: Raw AADT Data

Previous Research for 1995-2005

- n Growth curve models for different “groups” of interchanges to determine if *time* is statistically significant
- n Mixed effects models (location random, year fixed) for different “groups” of interchanges to determine if *location* is statistically significant

(Angela’s Thesis:

<http://theses.stat.ucla.edu/61/AngelaChangThesis.pdf>)

Previous research for 1973-2005

- n Fully imputed missing data using Functional Principal Component Analysis
- n Interpreted Principal Curves
- n Validated imputation results

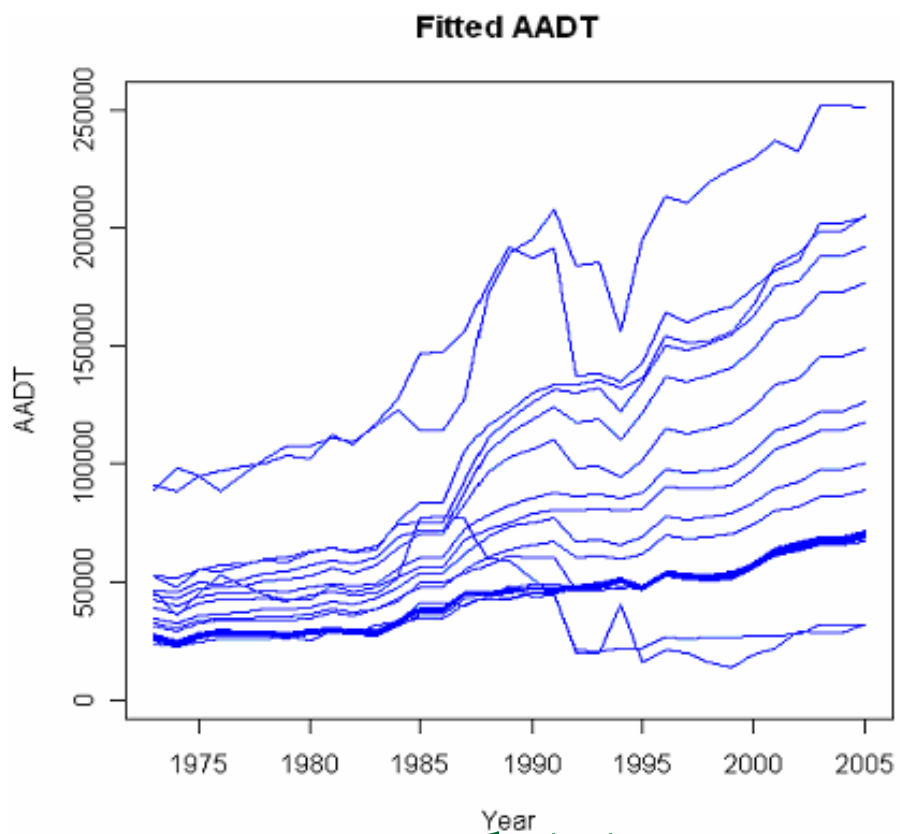


Figure 3: Fitted AADT Data

Further Research

- n Use imputed data to:
 - q Fit growth curve and mixed models
 - q Compare with results of Angela's thesis
 - q Forecast

Part A: Dataset Overview

Part 2: AADTT

Background

- n 4 intersections
- n North and South bound
- n Axle types 2-5
- n Year 1999 missing for all intersections
 - q Can be imputed with the average of 1998 and 2000 traffic



Figure 4: Raw AADTT Data

Further Research

- n Model truck traffic overall and by axle types
- n Incorporate port activity into models
- n Forecast

Part A: Dataset Overview

Part 3: Hourly Data Over 4 Weeks

Background

- n Location 1: Wheeler Ridge (09/29/06-10/28/06)
 - q Two datasets:
 - n North and South bound with total counts for each of four lanes in each direction
 - n North and South bound with counts for each of four lanes in each direction broken up by vehicle class (1-15)
- n Location 2: North of Junction 126 West (09/01/06-09/30/06)
 - q North and South bound with total counts for each of four lanes in each direction

Background (continued)

- n Location 3: Smokey Bear (09/01/06-09/30/06*)
 - q North and South bound with total counts for each of four lanes in each direction
 - n Includes counts for on-ramp

(*Until 11AM on 09/30/06.)

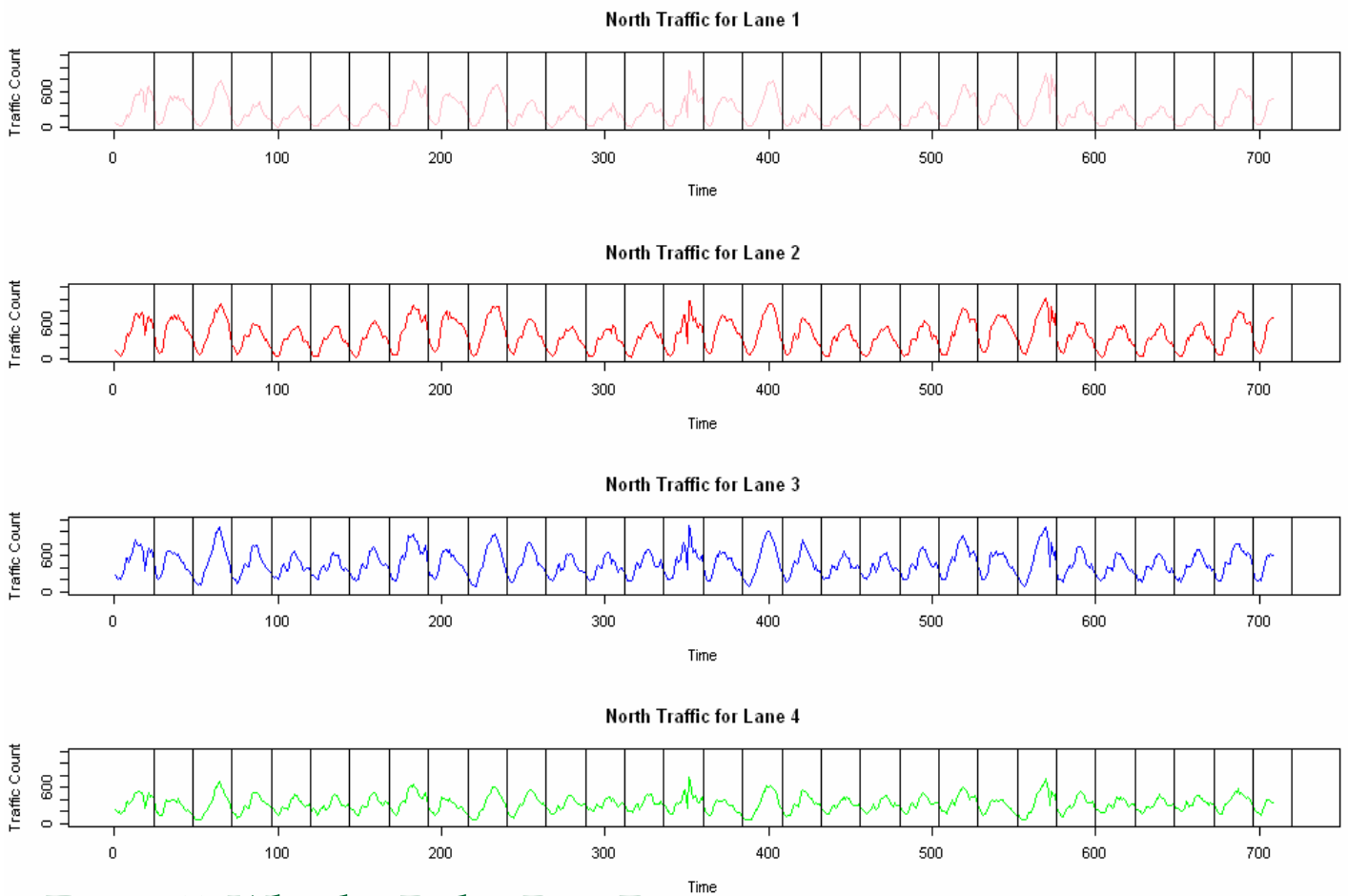


Figure 5: Wheeler Ridge Raw Data

Further Research

- n Model traffic by direction, and lane type
- n Incorporate day of week and time of day
- n Forecast

Part A: Dataset Overview

Part 4: Hourly Data Over 11 Years
(My research focus)

Background

- n North and South directions
 - n Total observations for each intersection:
96,432
 - q Corresponds to 4,018 days
 - n Missing data percentages:
 - q Route 126 - 8.82% (N) and 8.81% (S)
 - q Hungry Valley - 30.1%(N) and 33% (S)
 - q Wheeler Ridge - 29.9%(N) and 30% (S)
 - q Route 14 - 70.1%(N) and 61.1% (S)
-

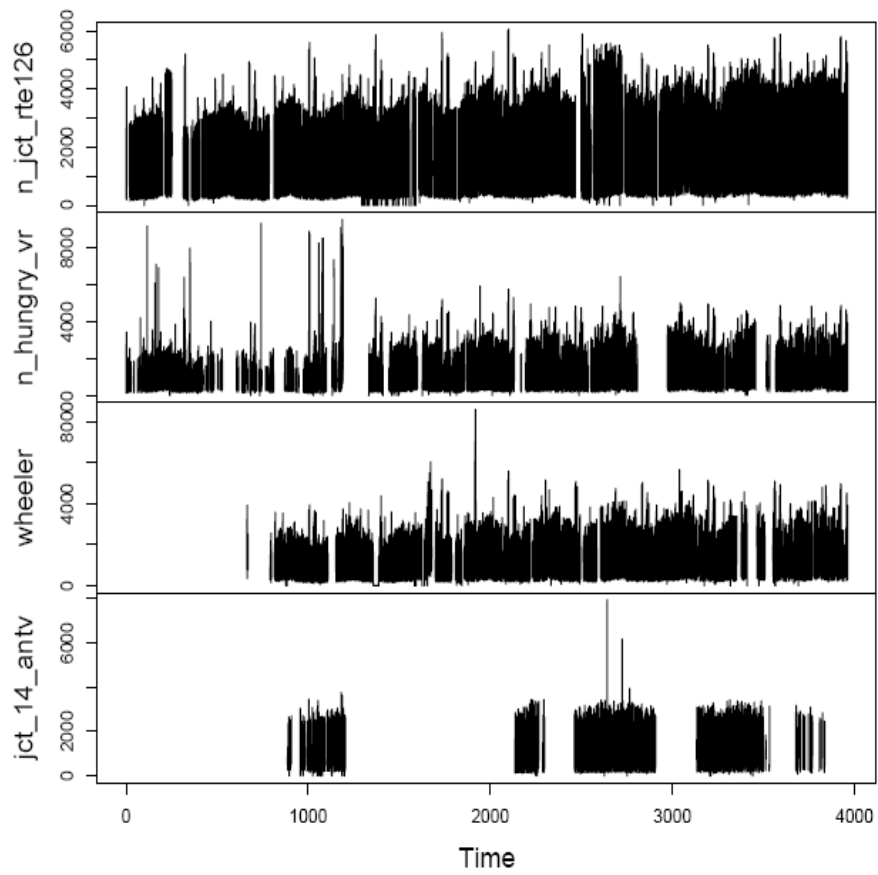


Figure 6: Raw North Counts

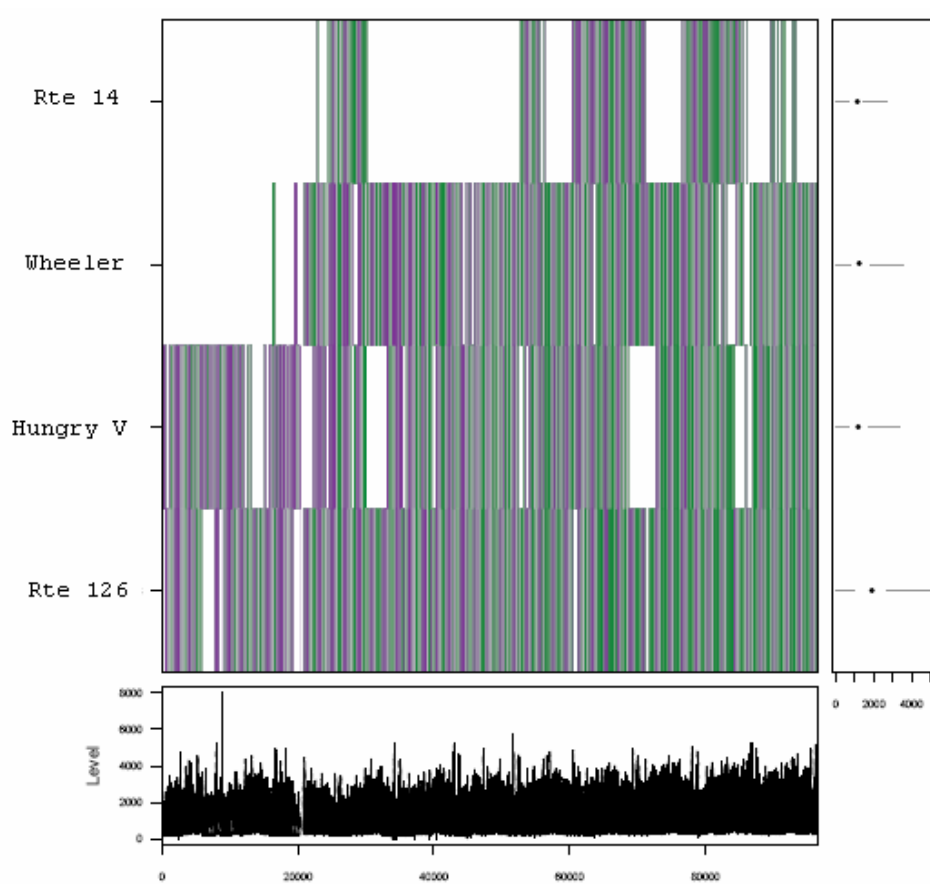


Figure 7: Raw North Counts: Multivariate TS Plot

Previous Research

- n Imputation of missing data via:
 - q Regression analysis
 - n Results: too smooth
 - n Residuals were not White Noise
 - q Functional Principal Component Analysis
- n Interpreted Principal Curves
- n Validated imputation results

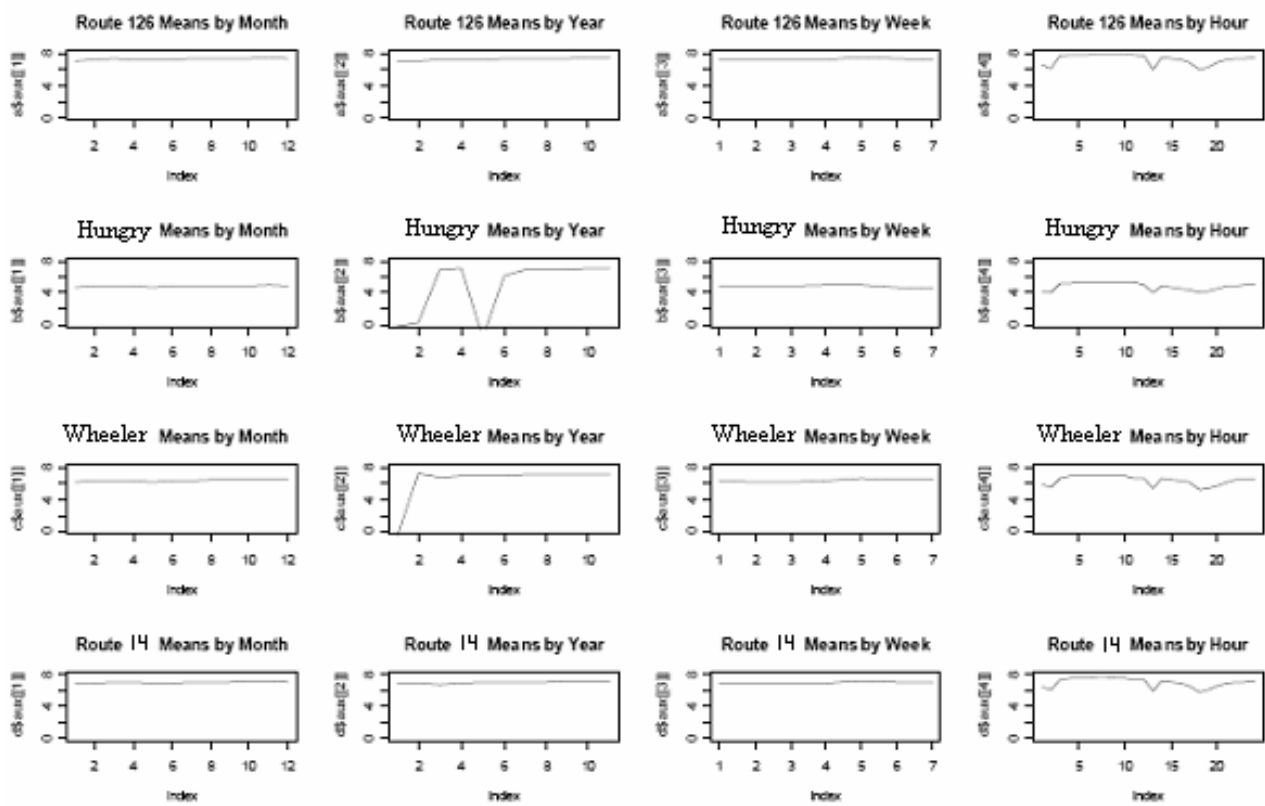


Figure 8: Imputation via Regression (North Direction) ²⁷

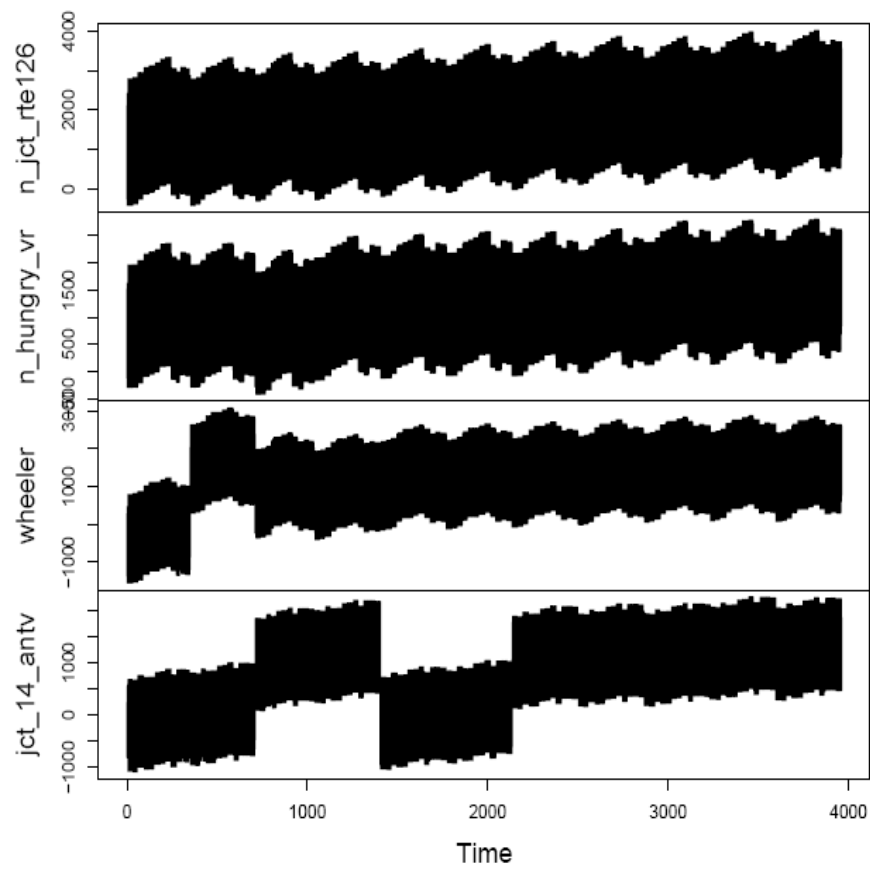


Figure 9: Imputation via Regression Results

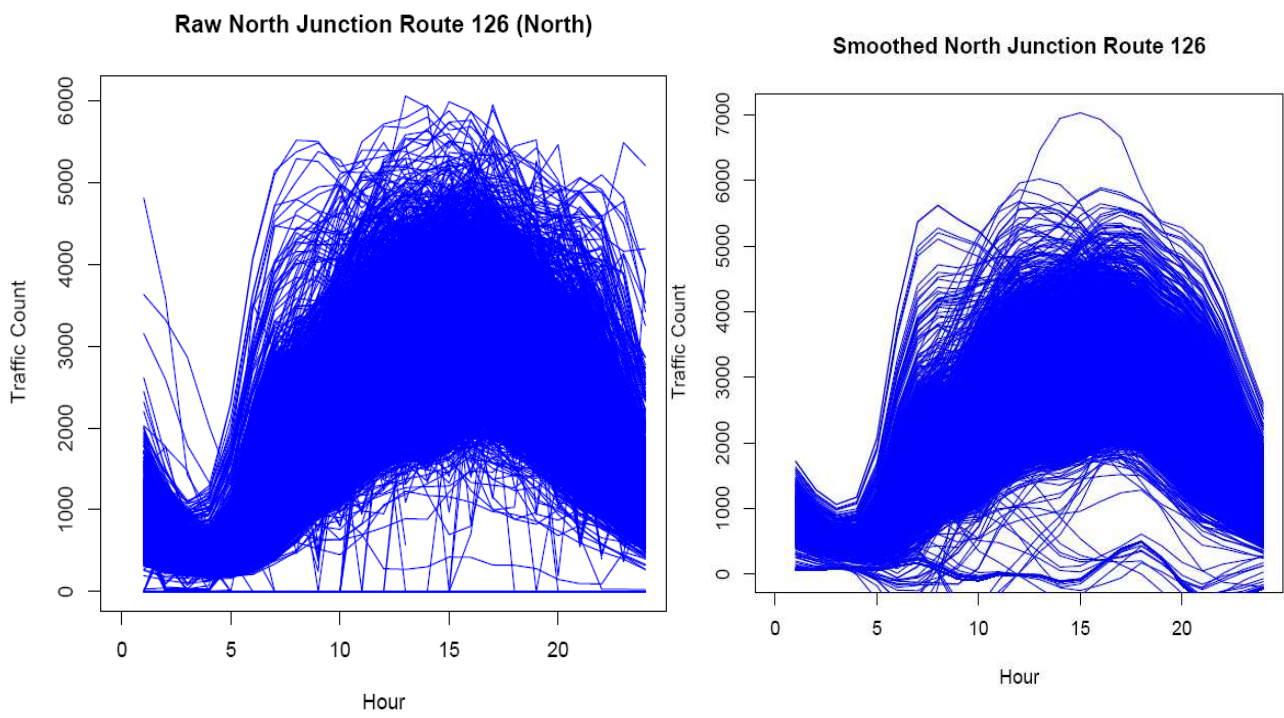


Figure 10: Before and After FPCA

North Junction Route 126: Actual vs. Fitted

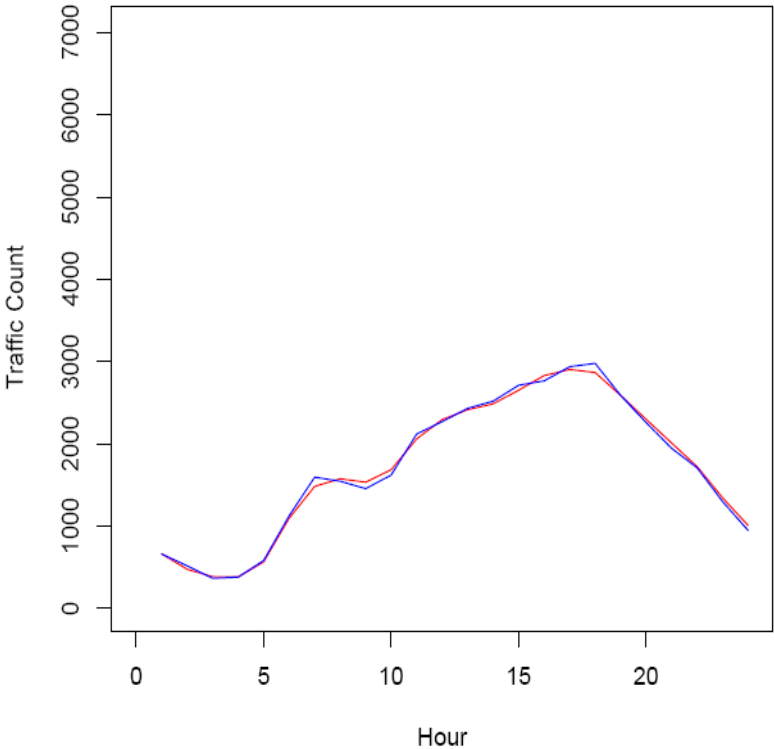


Figure 11: Comparison of Fit After FPCA is Applied
Legend: Original values are in Blue

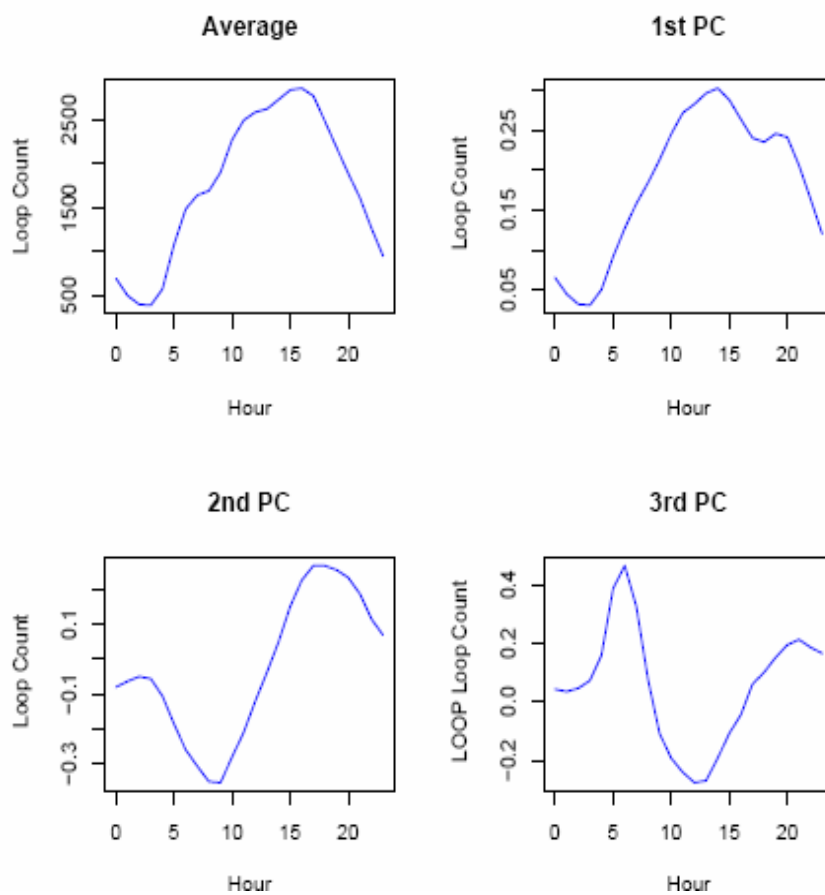


Figure 12: FPCA Principal Curves

Further Research

- n Research I would like to complete within the next 6-8 months:
 - q Combine LS and PCA:
 - n Perform Least Squares of the raw data
 - n Perform Principal Component Analysis on the residuals
 - q Time Series Analysis of Complete Data
 - q Forecast

Where To Go for Data?

- n Original data source for AADT and AADTT 1995-2006: <http://traffic-counts.dot.ca.gov/>
- n Formats available (for most datasets) through CES: XLS, TXT, RData
- n Email me (and let me know what dataset you would like to explore):
ikukuyeva@stat.ucla.edu

Any questions?

The speaker would like to thank Professor De Leeuw for resources and feedback, as well as Linda and Florence for help in getting the AADT and AADTT data into the current accessible formats.

Appendix

Overviews of Covered Topics

Vehicle Classification

Vehicle Class	Code	Description	Vehicle Weight (lbs.)
1	PC	Passenger cars	ALL
2	T1	Light-duty trucks	0 - 3,750
3	T2	Light-duty trucks	3,751 - 5,750
4	T3	Medium-duty trucks	5,751 - 8,500
5	T4	Light-heavy duty trucks	8,501 - 10,000
6	T5	Light-heavy duty trucks	10,001 - 14,000
7	T6	Medium-heavy duty trucks	14,001 - 33,000
8	T7	Heavy-heavy duty trucks	33,001 - 60,000
9	T8	Line-haul trucks	60,000 +
10	UB	Urban buses	ALL
11	MC	Motorcycles	ALL
12	SB	School buses	ALL
13	ME	Motor homes	ALL
14		Unclassified	ALL
15		Unclassifiable	ALL

Sources: www.fhwa.dot.gov/ohim/tmguides/word/tmg6.doc and o3.arb.ca.gov/msei/onroad/downloads/tsd/Vehicle_Population.doc

Principal Component Analysis (PCA)

- n May be used as a dimension-reduction technique:
 - q Suppose you have 200 variables (and you do not know where to start analyzing them)
 - q PCA will use linear combinations of the original variables to explain the (same amount of) variation found in the dataset with fewer variables
 - n Typically, 2-3 will now explain 80% of the variation of the dataset

(Please see Stats. 101B or 202B for more information.)

Functional Principal Component Analysis (FPCA)

n *Idea:*

q FPCA estimates curves that represent our discrete data

n In this case, one for each day:

q 4,018 curves in all (for my research)

q In the construction of the curves, FPCA uses information from data points of the same day (9)

(Please see *Curves* paper for more information.)

FPCA: Handling missing data

- n *Step 1:* Perform Principal Component Analysis on the observed data
- n *Step 2:* Estimate the curves
 - q In our case, one for each day
- n *Step 3:* If a particular day is missing use the estimated curve for that day
- n *Step 4:* Perform Principal Component Analysis on the complete data
- n Repeat Steps 2-4 until convergence (33)

How to Find the Curves Paper?

- n Visit Professor De Leeuw's page:
 - q <http://idisk.mac.com/jdeleeuw-Public?view=web>
 - n Folder: *curves*