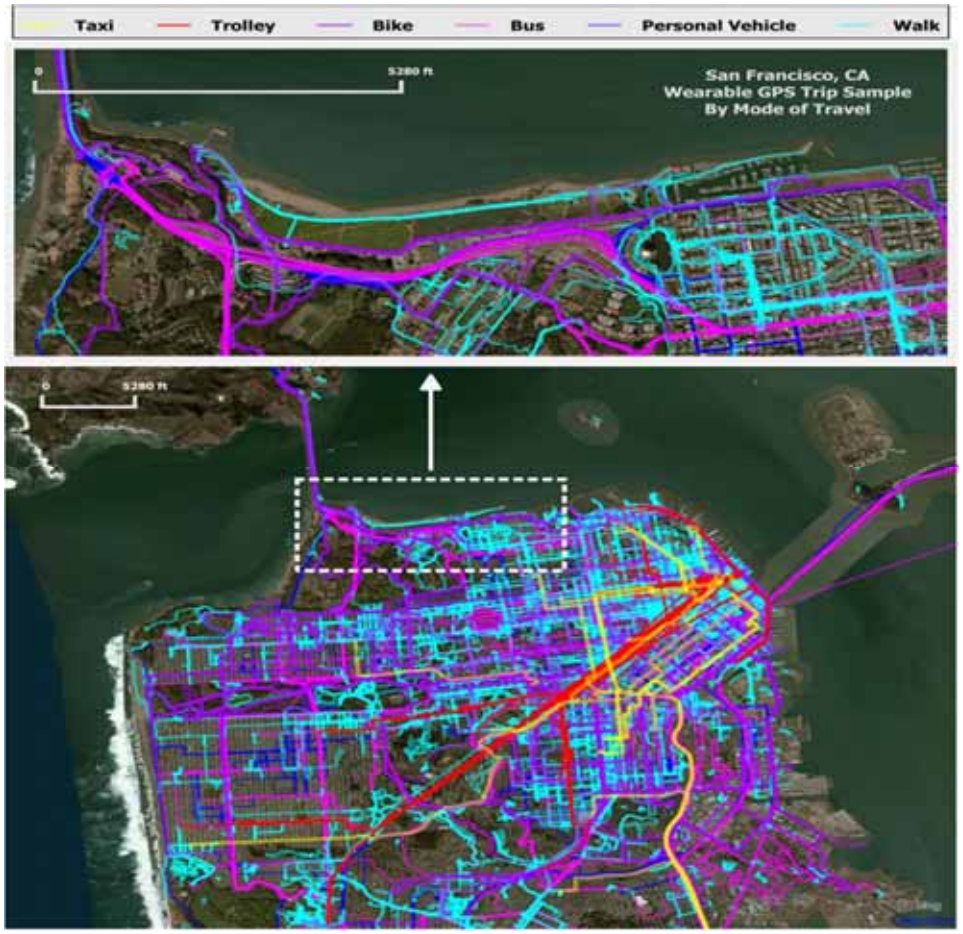
Transportation Secure Data Center at NREL

NATIONAL RENEWABLE ENERGY LABORATORY



USDOT Datapalooza
June 17, 2015

Elaine Murakami
Elaine.murakami@dot.gov
FHWA Office of Planning

Goals: Protect, Access, and Preserve

Data at risk of disappearing
Risk of disclosure of individual responses

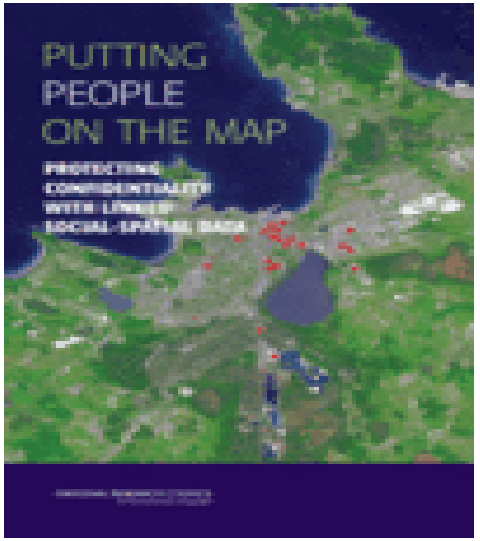
Extend and expand the use of data collected with limited research funds

Curation, from the Latin word “to care”

Transportation Secure Data Center (TSDC) Rationale

Secure data center makes data available for legitimate research while preserving privacy

- Maximizes value from limited public funds
- Benefits data providers and users
  - Takes care of archiving and responding to data requests
  - Data accessible from a central location



* See this 2007 National Research Council report: **http://books.nap.edu/openbook.php?record_id=11865**

Why choose NREL?

NREL Transportation Data Centers in addition to the TSDC

Alternative Fuels Data Center (AFDC)

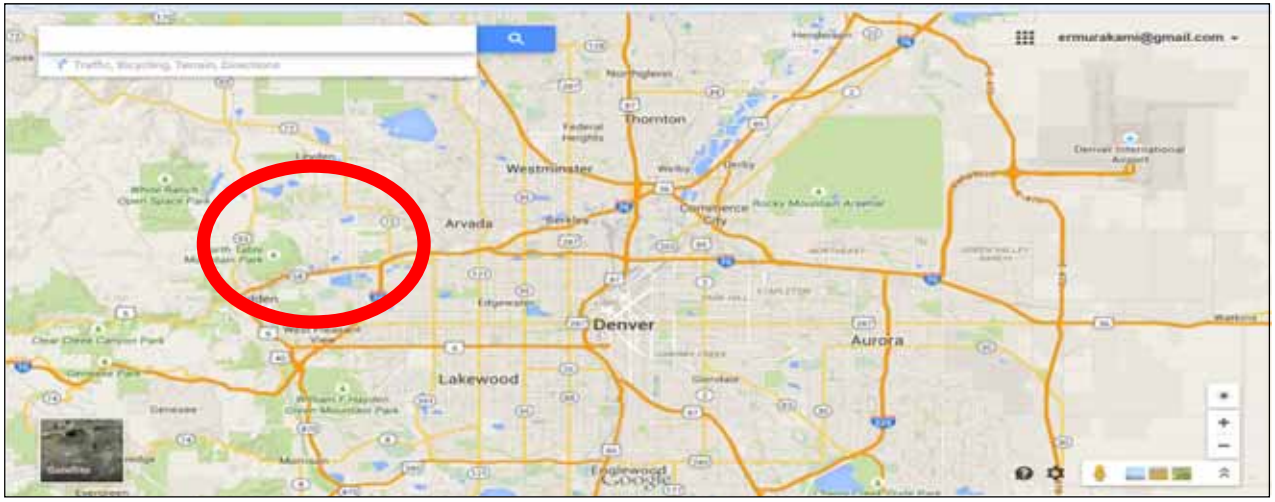
Public clearinghouse of information on the full range of advanced vehicles and fuels

National Fuel Cell Technology Evaluation Center (NFCTEC)

Industry data and reports on hydrogen fuel cell technology status, progress, and challenges

Medium- and heavy-duty drive-cycle and powertrain data from advanced commercial fleets

FleetDASH: *Business intelligence to manage Federal fleet petroleum/alternative fuel consumption*



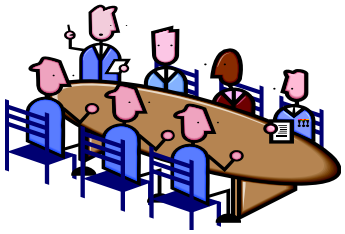
Example Datasets

Dataset	# Vehicles	# Days	# Persons	# Days
2013 Mid-Region Council of Governments (Albuquerque, NM)			931	3
2010-2012 California Statewide Household Travel Survey	3,910	7	7,574	3
2011 Atlanta Regional Household Travel Survey	1,653	7	797	3
2010 Metropolitan Council (Minneapolis/St Paul, MN) T			174	7
2007 Chicago Regional Household Travel Inventory	408	7	209	7
2004-2006 Puget Sound Traffic Choices Study	484	540		
2004 Mid-America Regional Council (Kansas City) Regional Travel Study	408	5		

Developing the TSDC Operating Procedures

Maintain balanced focus on dual priorities

- Privacy protection
- Maximize usability



An advisory committee provides oversight

- Group includes data providers and users
- Represents industry, academia and government

Reference best practice

- other NREL data centers
- U.S. Census Research Data Center program; virtual data centers on social science[1] and Medicare/Medicaid data[2]
- National Household Travel Survey (NHTS)

TSDC Data Archiving Procedures

- Establish MOU agreement with data provider
  - Receive data via mail or secure FTP

- Load onto secure raw data handling server
  - Restricted access
  - On-site security force
  - Established cyber security group

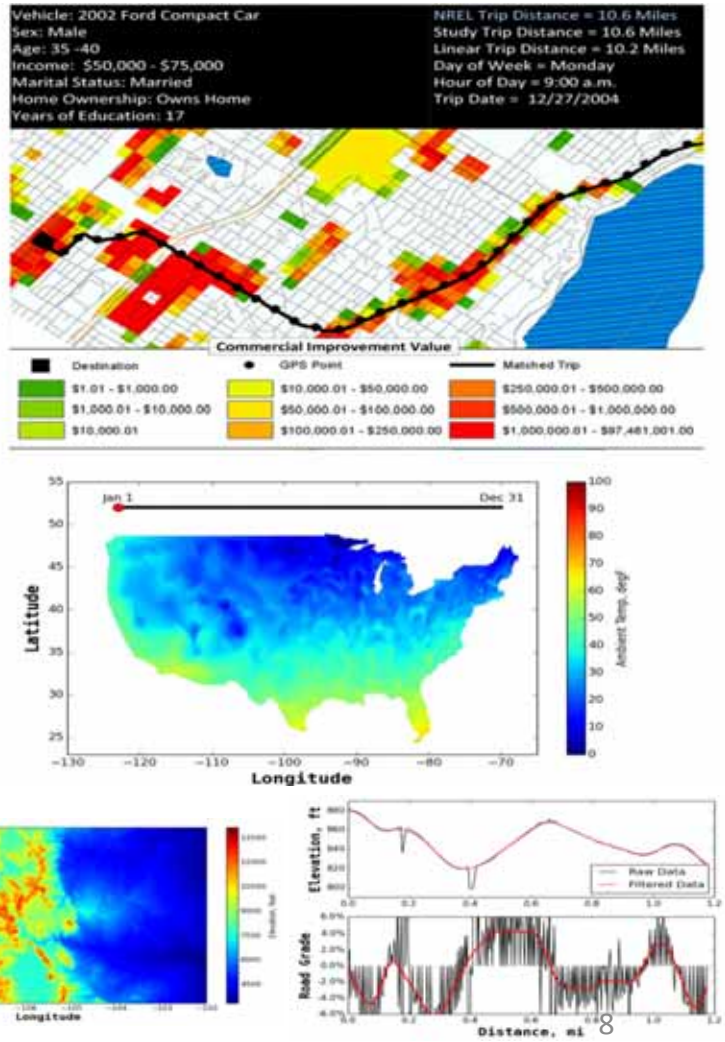
- Maintain data backups
  - Data mirrored on large storage array
  - Maintain backup in separate location for fire/disaster protection

NREL Data Center storage arrays

MOU = memorandum of understanding; FTP = file transfer protocol

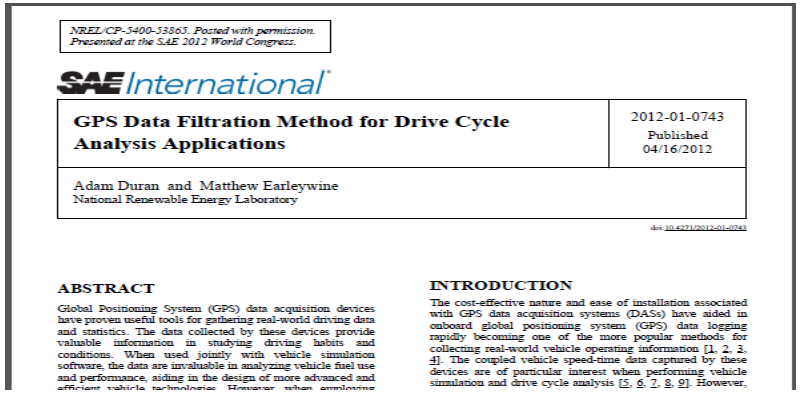
TSDC Data Processing

- Standardize formatting
  - Raw point lat/long, timestamp, precision
  - Trip-level distance and time summary
  - Household/vehicle demographic information
- Remove explicitly identifying information
  - Participant names, addresses, contact info
- Quality control for errant/missing GPS points
  - Remove, adjust and/or interpolate points
  - Maintain in both processed (filtered) and original (raw/uncorrected) formats
- Add/link to reference data
  - Road network, road grade, GIS layers
  - Meteorological, economic, land use data
  - Vehicle and demographic information



Details on GPS Data Filtration

1. **Remove duplicate records and data with negative values or differential time steps**
2. **Replace outlying high/low speed values**
3. **Remove zero-speed signal drift when vehicle is stopped**
4. **Replace false zero-speed records**
5. **Amend gaps in data**
6. **Repair outlying acceleration/deceleration values**
7. **Denoise and condition final signal**



NREL/CP-5400-53863. Posted with permission.
Presented at the SAE 2012 World Congress.

SAE International

GPS Data Filtration Method for Drive Cycle Analysis Applications

2012-01-0743
Published
04/16/2012

Adam Duran and Matthew Earleywine
National Renewable Energy Laboratory

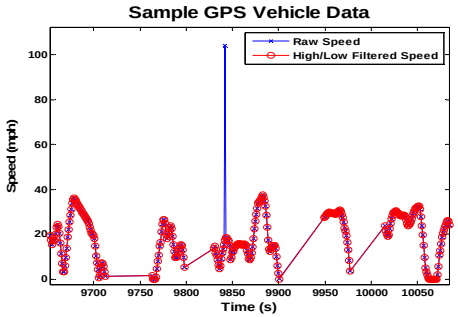
doi: 10.4271/2012-01-0743

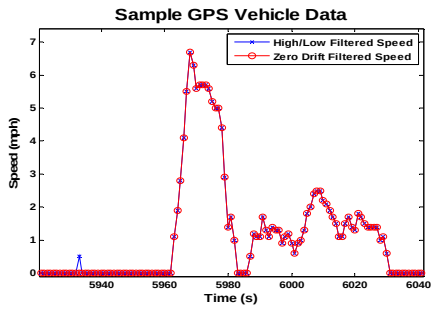
ABSTRACT

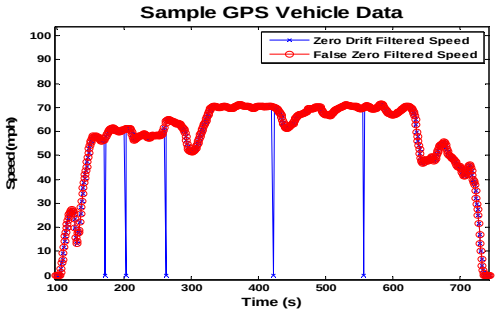
Global Positioning System (GPS) data acquisition devices have proven useful tools for gathering real-world driving data and statistics. The data collected by these devices provide valuable information in studying driving habits and conditions. When used jointly with vehicle simulation software, the data are invaluable in analyzing vehicle fuel use and performance, aiding in the design of more advanced and efficient vehicle technologies. However, when employing

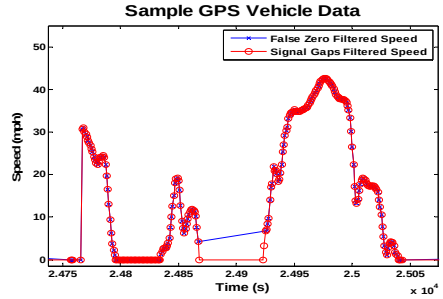
INTRODUCTION

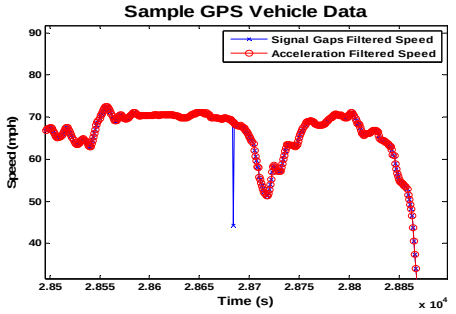
The cost-effective nature and ease of installation associated with GPS data acquisition systems (DASs) have aided in onboard global positioning system (GPS) data logging rapidly becoming one of the more popular methods for collecting real-world vehicle operating information [1, 2, 3, 4]. The coupled vehicle speed-time data captured by these devices are of particular interest when performing vehicle simulation and drive cycle analysis [5, 6, 7, 8, 9]. However,

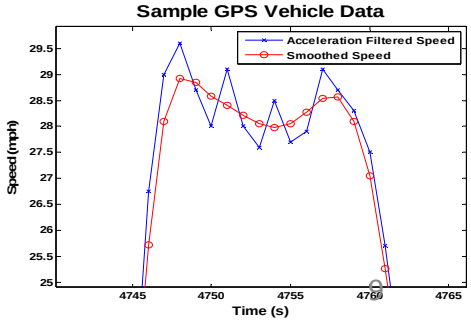




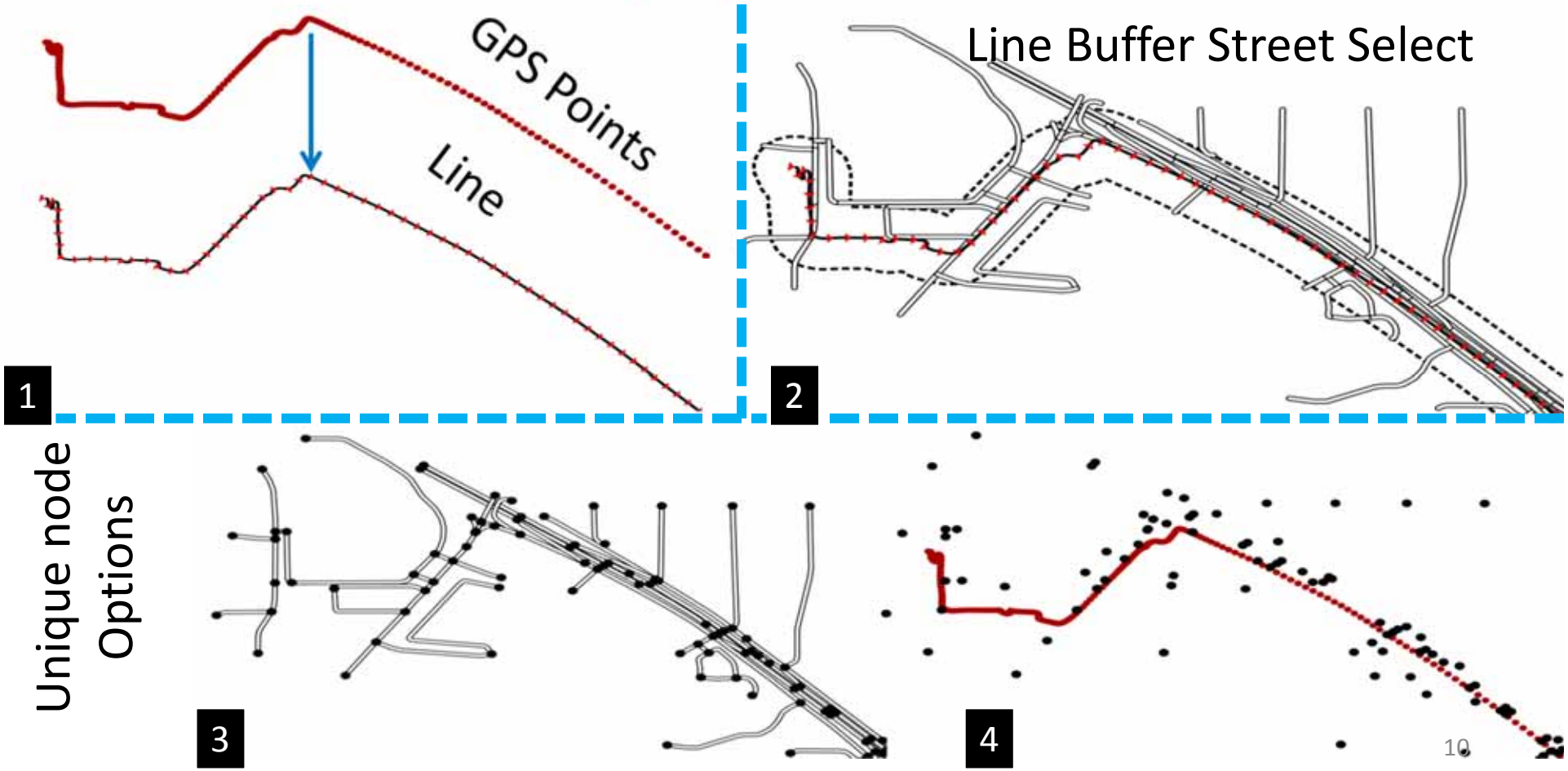






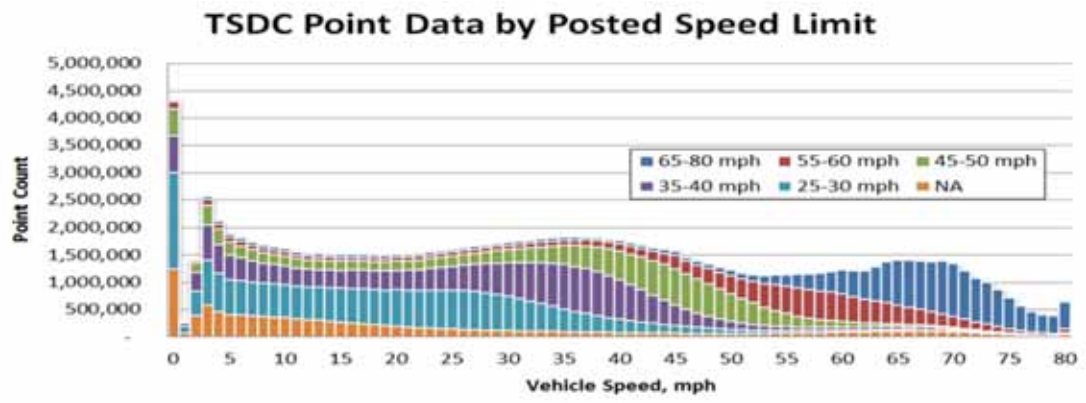


Map Matching: Dynamic node selection

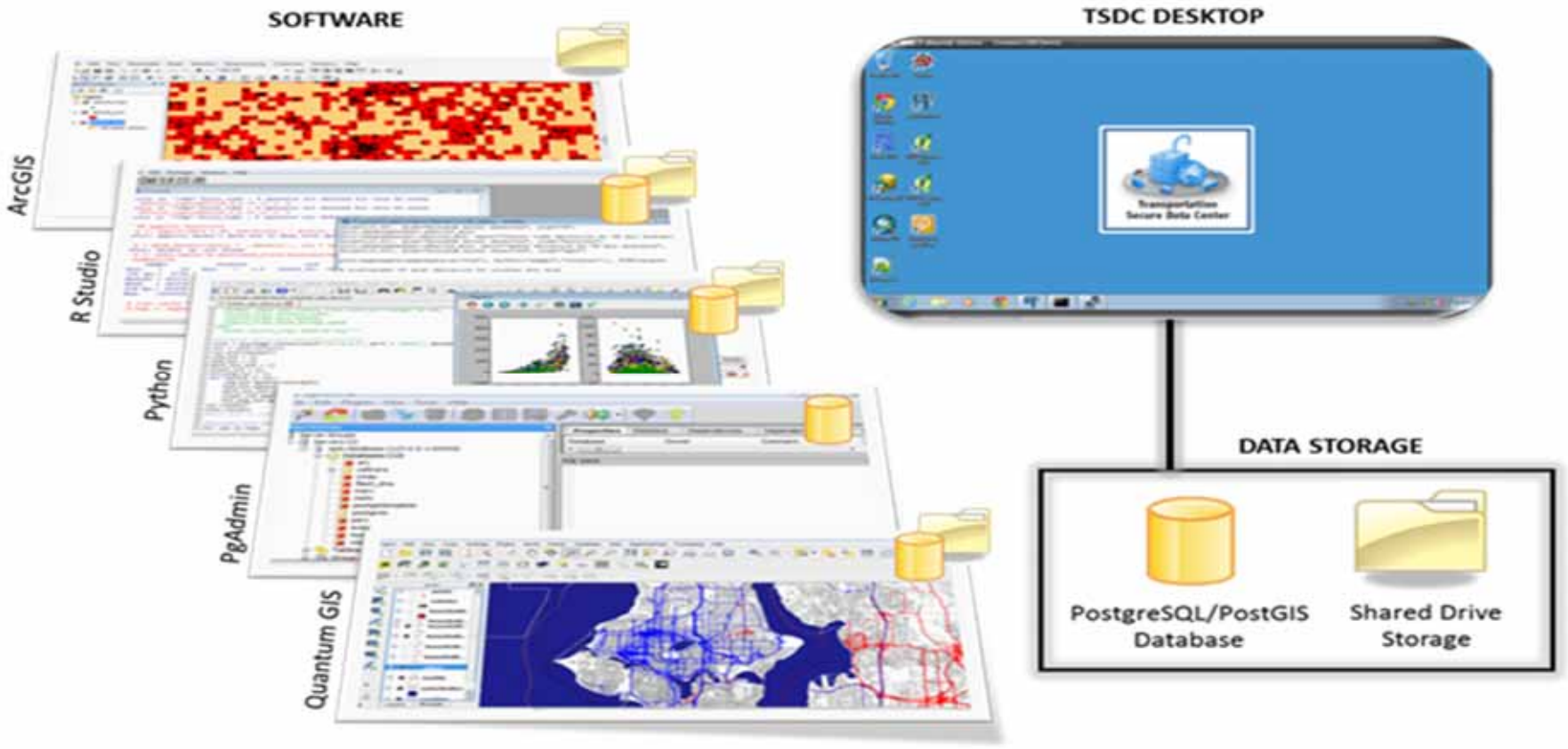


Examples of types of analyses

- Safety: driver volatility (accelerations and declerations)
- Air quality and emissions
- Electric vehicle and battery design
- Activity-based microsimulation of travel
  - for long range forecasting
  - For toll/pricing studies
  - Mode choice modeling
- Parking search (new)
- VMT estimates with vehicle types, land use, or hhld income
- Electrification of the roadway



TSDC Secure Portal Snapshot



Questions?

For More Information on the TSDC…

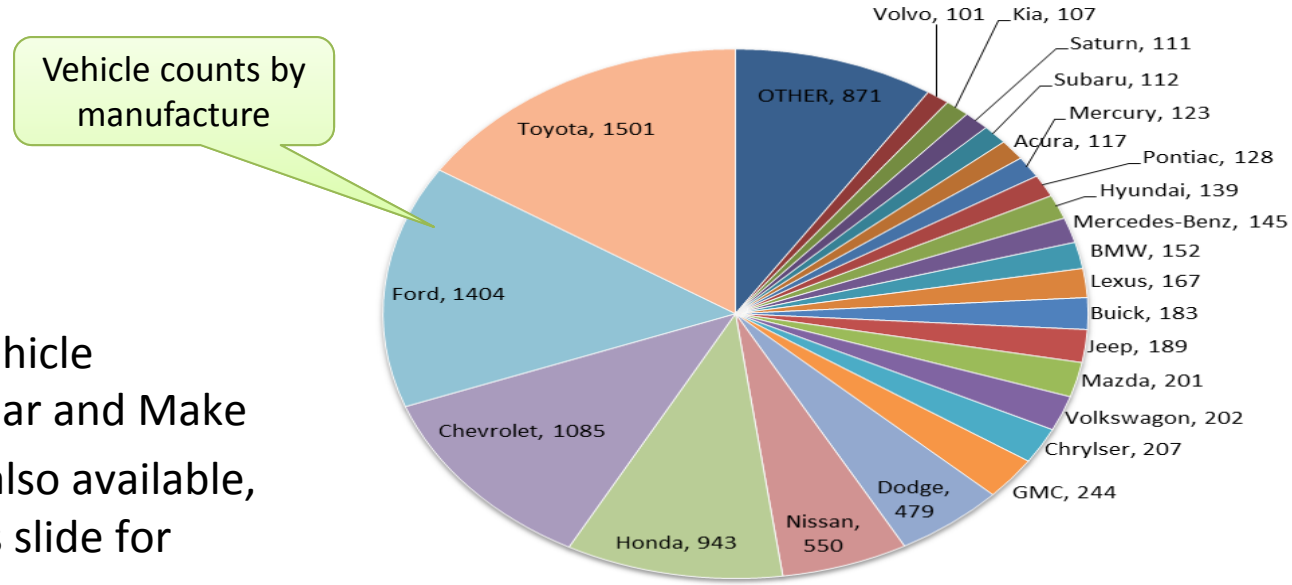
Visit the website: www.nrel.gov/tsdc

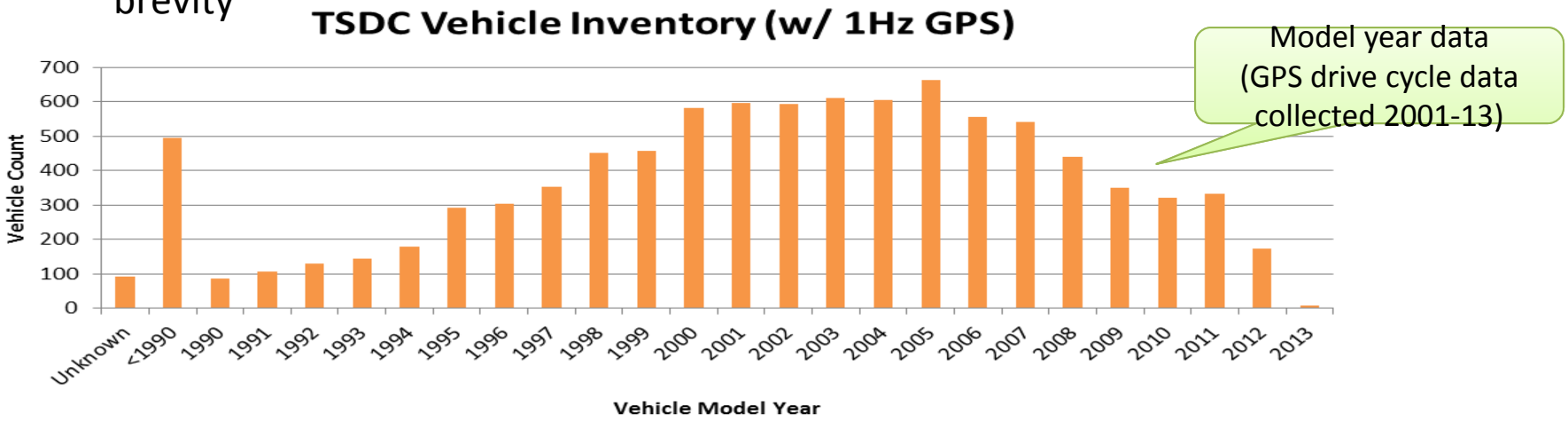
- Read about the project
- View fact sheets and publications
- Download cleansed public data
- Apply for secure portal access
- Sign up to receive e-mail updates

Contact: Jeff.Gonder@nrel.gov or tsdc@nrel.gov

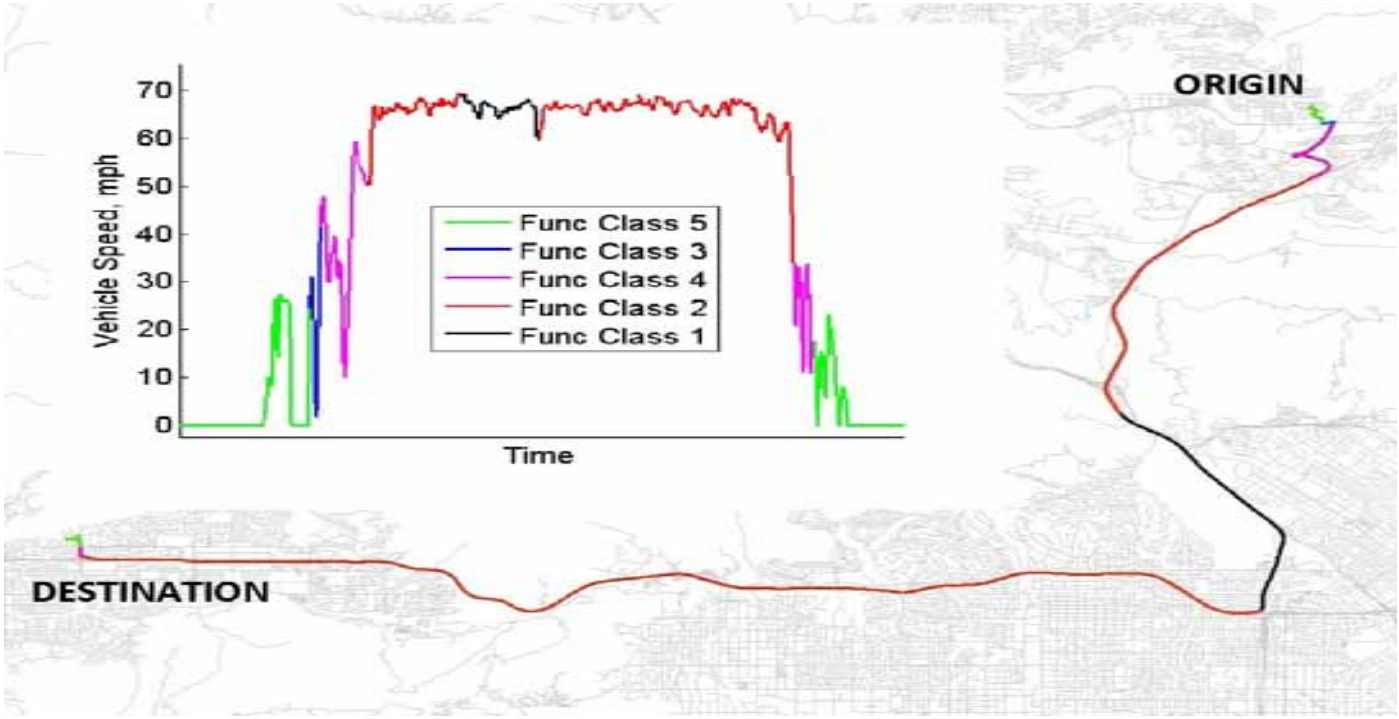
Examples of data analyses

- Breakdown of TSDC vehicle inventory by Model Year and Make
- Model information is also available, but excluded from this slide for brevity

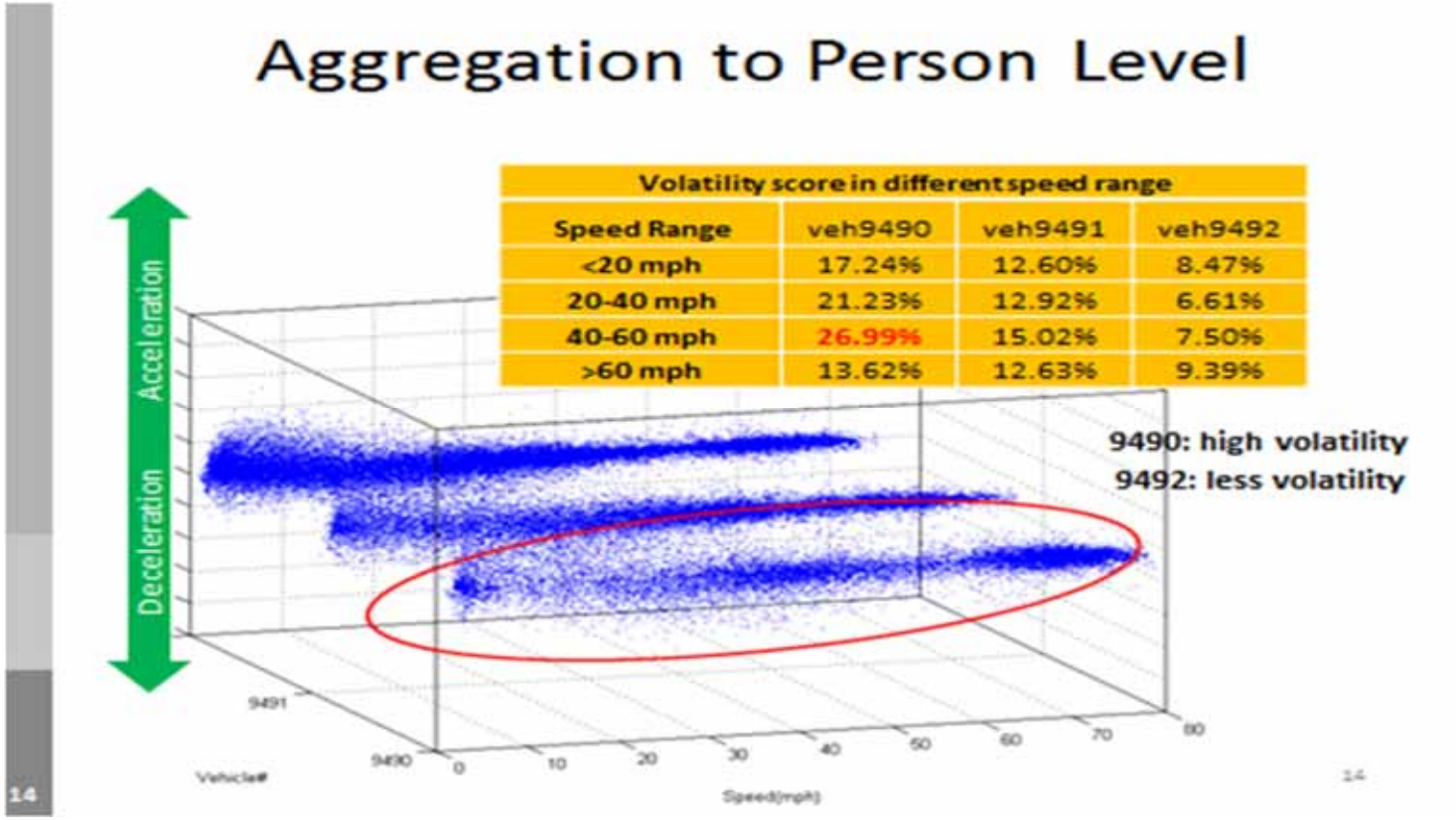




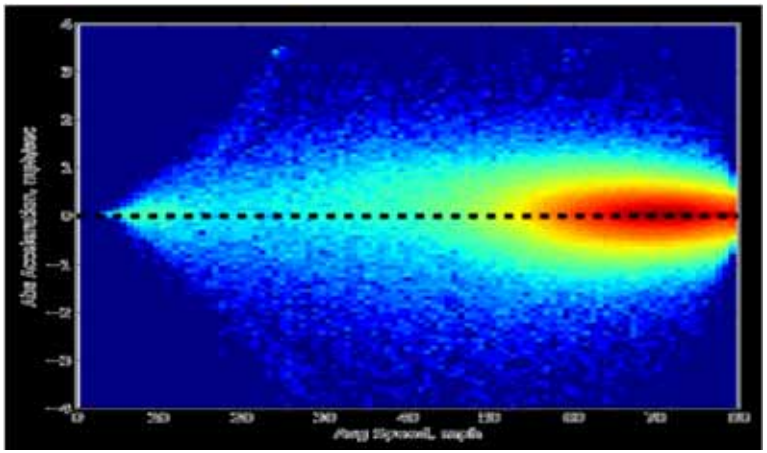
Speed profiles by Roadway Func Class



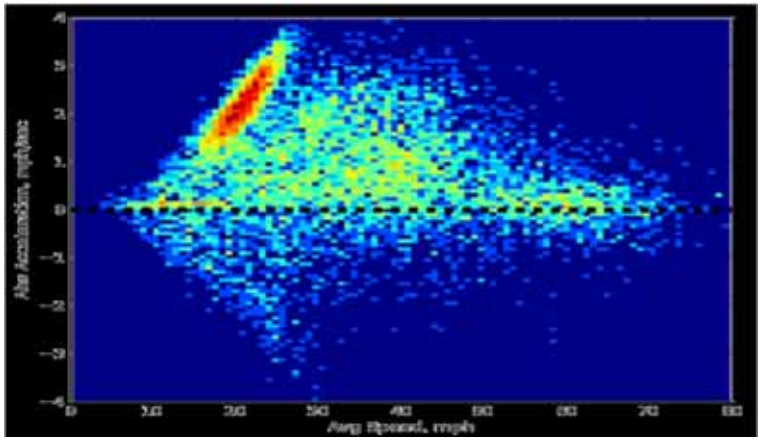
Asak Khatak et al on jerking



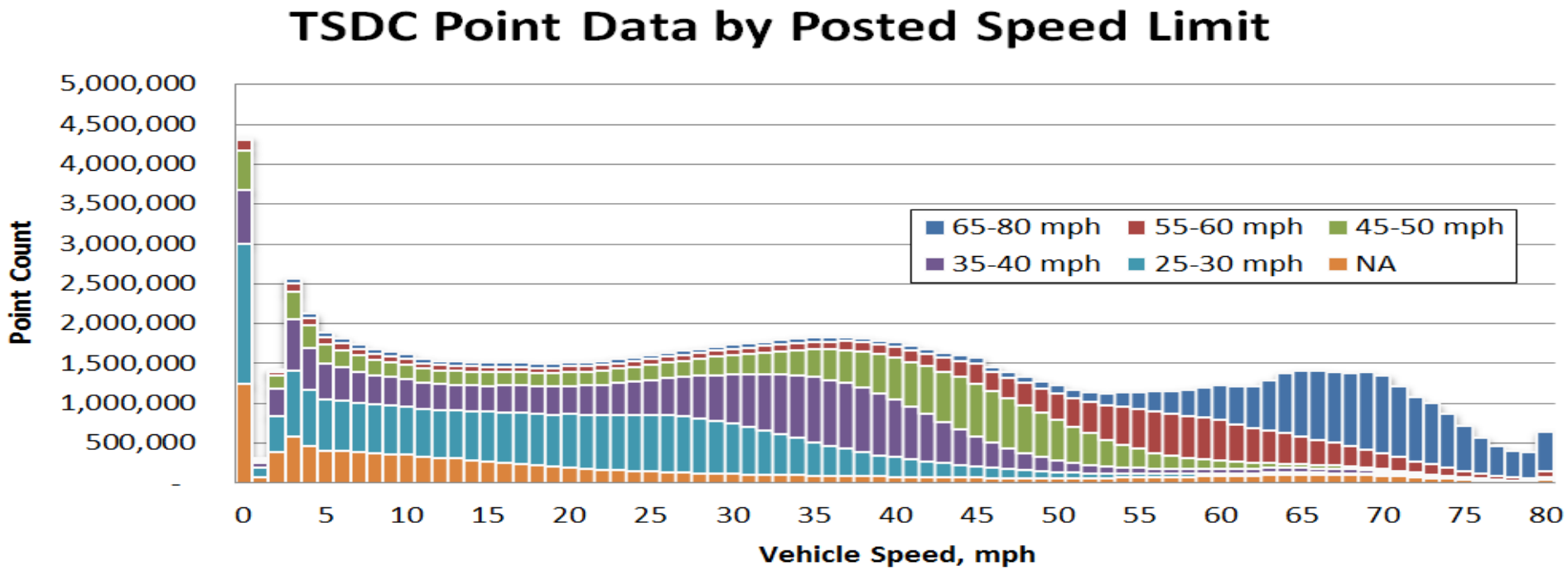
Speed and accelerations by roadway functional classification



Func class = 1



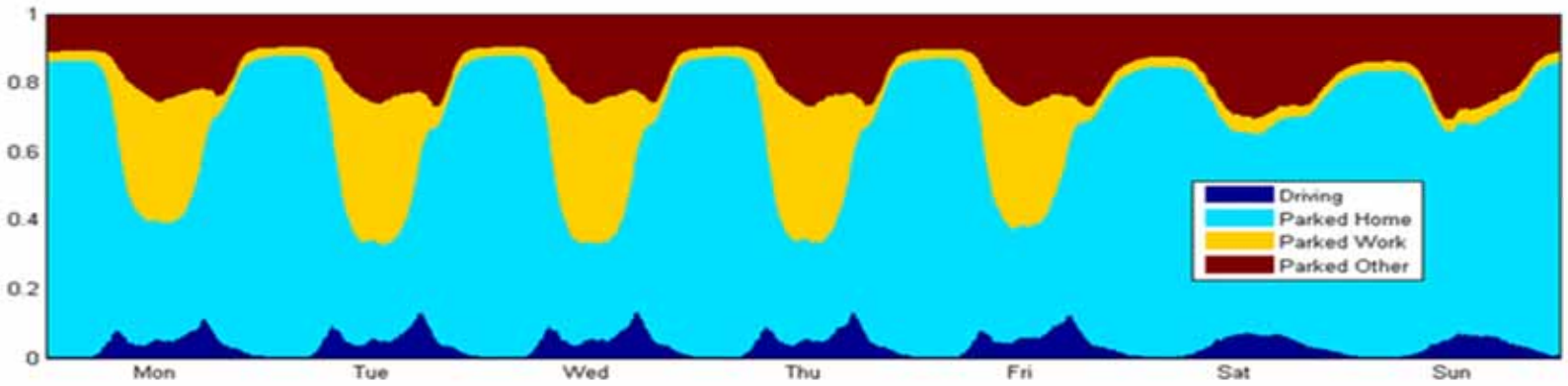
Func class = 3, after leaving a func class 4 roadway



- 31M 1Hz GPS points in TSDC
- Histogram of vehicle speed data reveals local maxima near 0, 37, and 67 mph
- Recorded driving speeds generally agree with posted speed limits

Transportation Secure Data Center

Driving Variability by Day of Week



Other studies of variability:

- Travel time reliability by roadway types
- Cluster analysis of travelers, using multiday analysis and dispersion of destinations
- Mode choice variation, esp driving and transit users