



Overview of the Transportation Secure Data Center (www.nrel.gov/tsdc)



November 2015

**Jeff Gonder – Senior Engineer/Supervisor and TSDC Project Leader
National Renewable Energy Laboratory (NREL) Transportation Center**

Transportation Secure Data Center (TSDC) Rationale

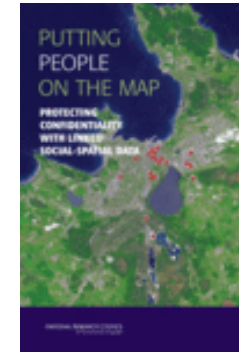


High-resolution survey data (e.g., GPS travel profiles, geo-coded trip ends)

- Very valuable for research
- Misuse could violate participant privacy

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



The TSDC has been **supported since 2009 by NREL, U.S. DOT and U.S. DOE**

- Department of Transportation, Federal Highway Administration
- Department of Energy, Vehicle Technologies Office

NREL Transportation Data Centers

Secure Access, Expert Analysis and Validation Support Decision-Making

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

Transportation Secure Data Center (TSDC): *Detailed fleet data, including GPS travel profiles*

Fleet DNA Data Collection

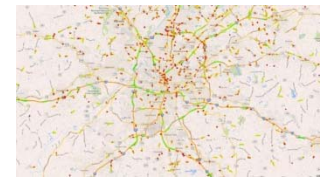
Medium- and heavy-duty drive-cycle and powertrain data from advanced commercial fleets

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

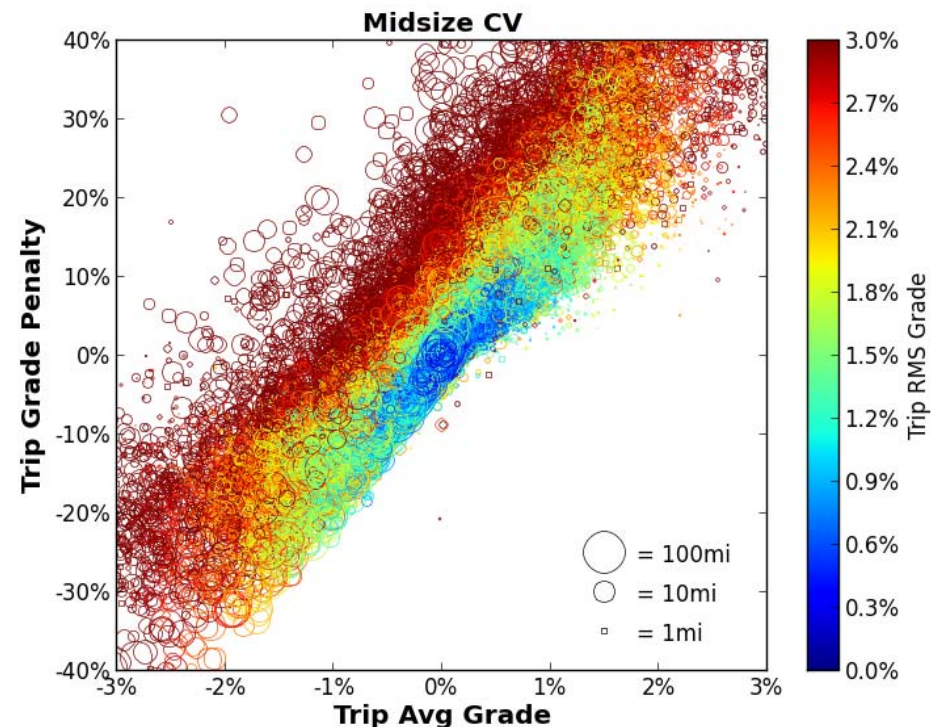
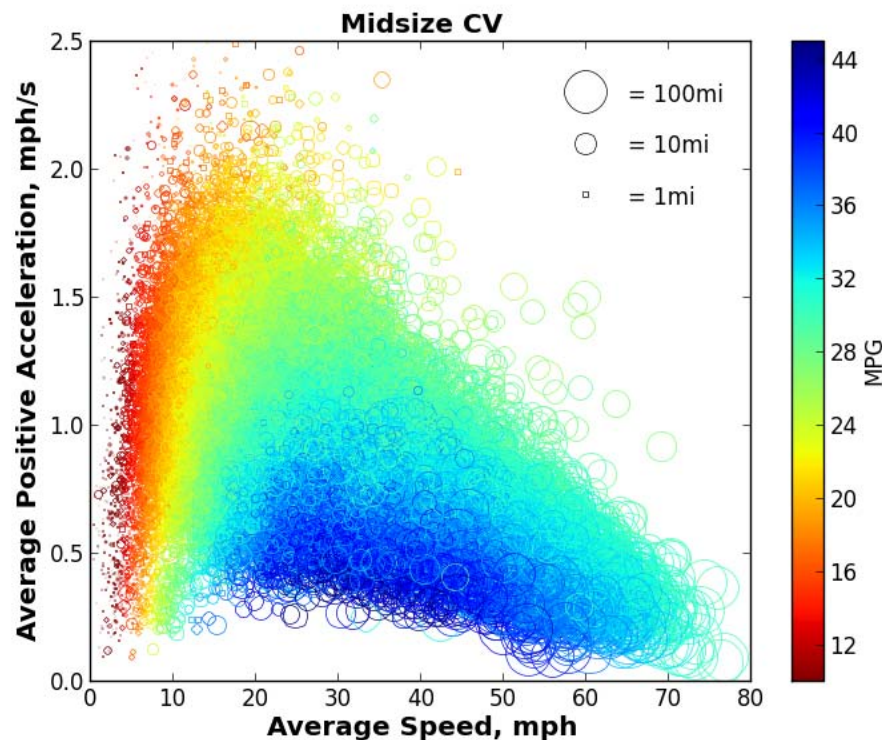
Features	AFDC	NFCTEC	TSDC	Fleet DNA	Fleet DASH
Securely Archived Sensitive Data		Y	Y	Y	Y
Publicly Available Cleansed Composite Data	Y	Y	Y	Y	
Quality Control Processing	Y	Y	Y	Y	Y
Spatial Mapping/GIS Analysis	Y	Y	Y	Y	Y
Custom Reports		Y		Y	Y
Controlled Access via Application Process			Y		
Detailed GPS Drive-Cycle Analysis			Y	Y	

Related Real-World Analysis Efforts Using TSDC Data

Large distribution of real-world GPS travel profiles, including speed, acceleration, distance, time of day, stop duration, etc. E.g., previous analysis explored fuel economy sensitivity to speed/acceleration characteristics and road grade using hundreds of thousands of GPS drive cycles in NREL TSDC

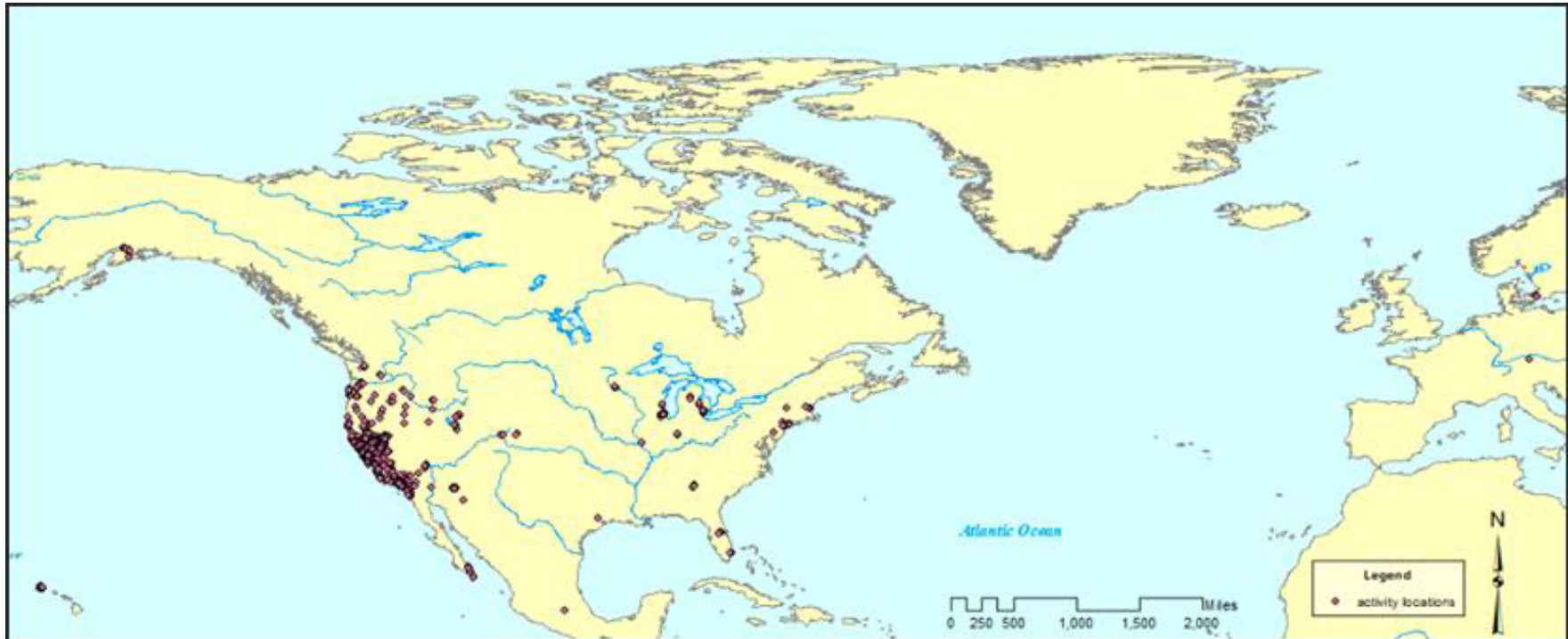


Data Visual



GPS = Global Positioning System; CV = Conventional Vehicle

Example Travel Behavior Analysis: Day-to-Day Destination Variation for CA Bay Area Travelers



Consider short- and long-distance work commutes and leisure travel
Able to clearly distinguish patterns of variability in terms of number of trips and type and dispersion of destinations

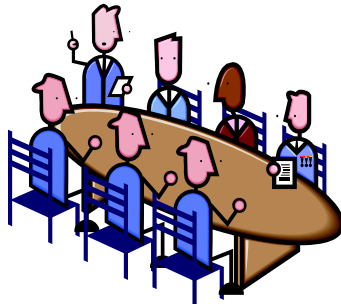
K. Deutsch-Burgner. "Multiday Variation in Time Use and Destination Choice in the Bay Area Using the California Household Travel Survey." Report on Multiday GPS Travel Behavior Data for Travel Analysis (2015).

http://www.fhwa.dot.gov/planning/tmip/publications/other_reports/multiday_gps/fhwahep15026.pdf

Developing the TSDC Operating Procedures

Maintain **balanced focus** on dual priorities

- Privacy protection first and foremost
- Maximize usability (within constraints)



An **advisory committee** helps support oversight

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

Reference best practice examples

- Experience from other NREL data centers
- And examples external to NREL (e.g., U.S. Census Research Data Center program; virtual data centers on social science¹ and Medicare/Medicaid data²)



¹ - www.dataenclave.org; ² - www.resdac.org/cms-data/request/cms-virtual-research-data-center

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

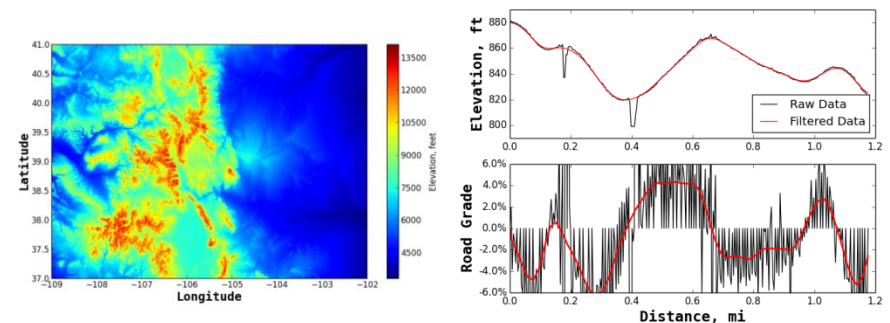
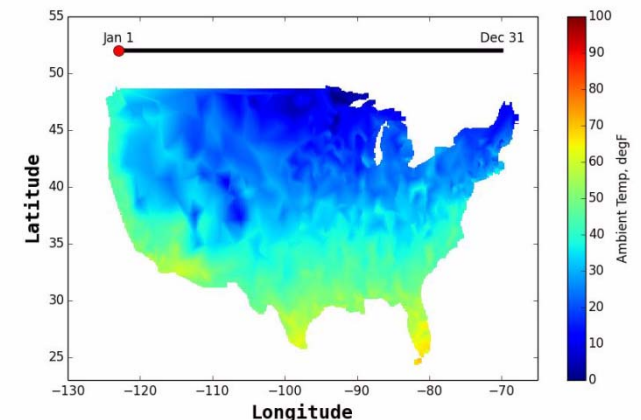
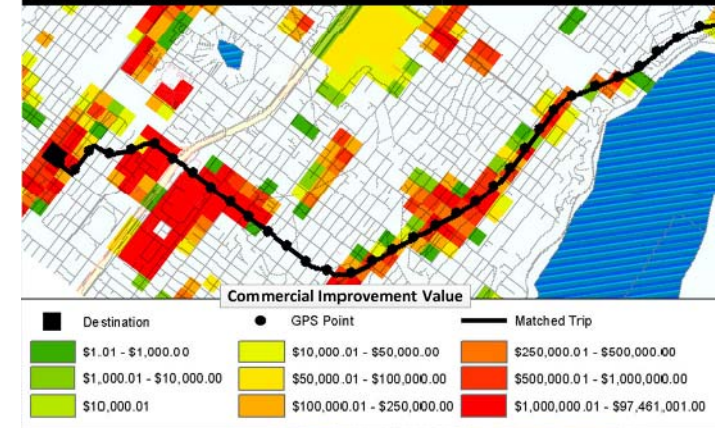
MOU = memorandum of understanding; FTP = file transfer protocol

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

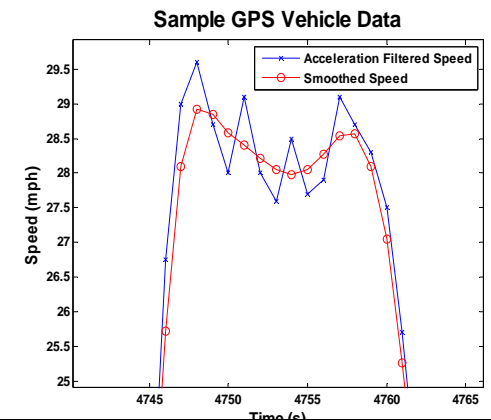
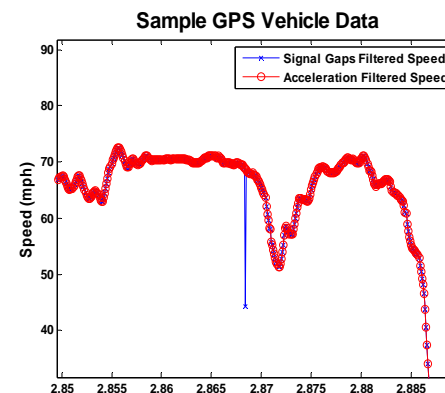
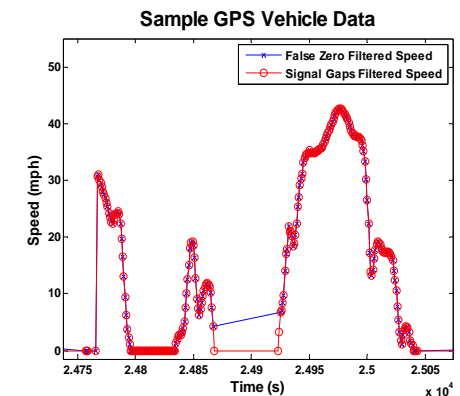
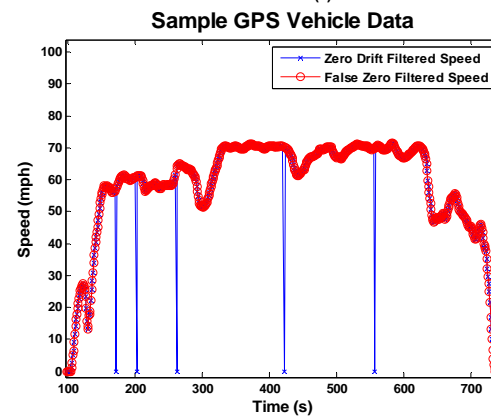
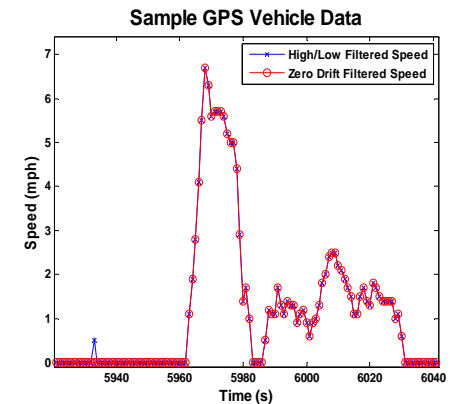
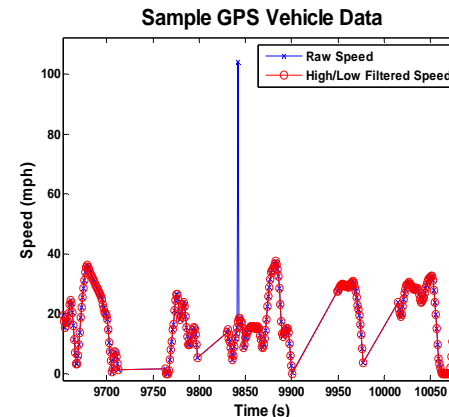
Vehicle: 2002 Ford Compact Car
 Sex: Male
 Age: 35 -40
 Income: \$50,000 - \$75,000
 Marital Status: Married
 Home Ownership: Owns Home
 Years of Education: 17

NREL Trip Distance = 10.6 Miles
 Study Trip Distance = 10.6 Miles
 Linear Trip Distance = 10.2 Miles
 Day of Week = Monday
 Hour of Day = 9:00 a.m.
 Trip Date = 12/27/2004



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-5400-53865. 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 analyzing

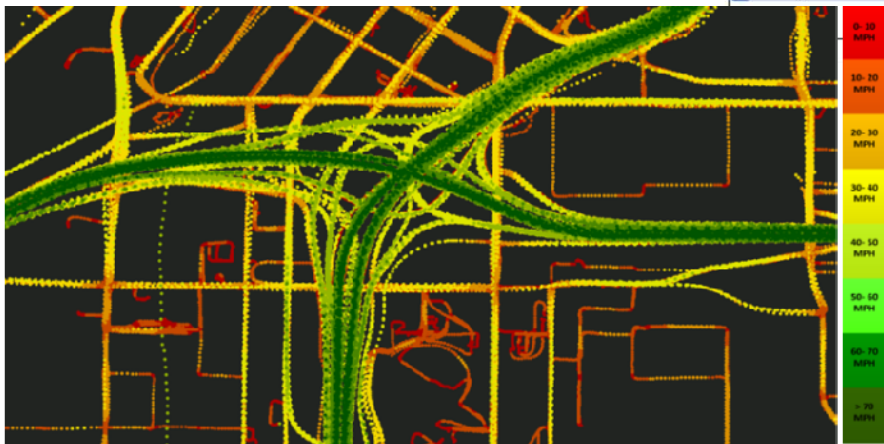
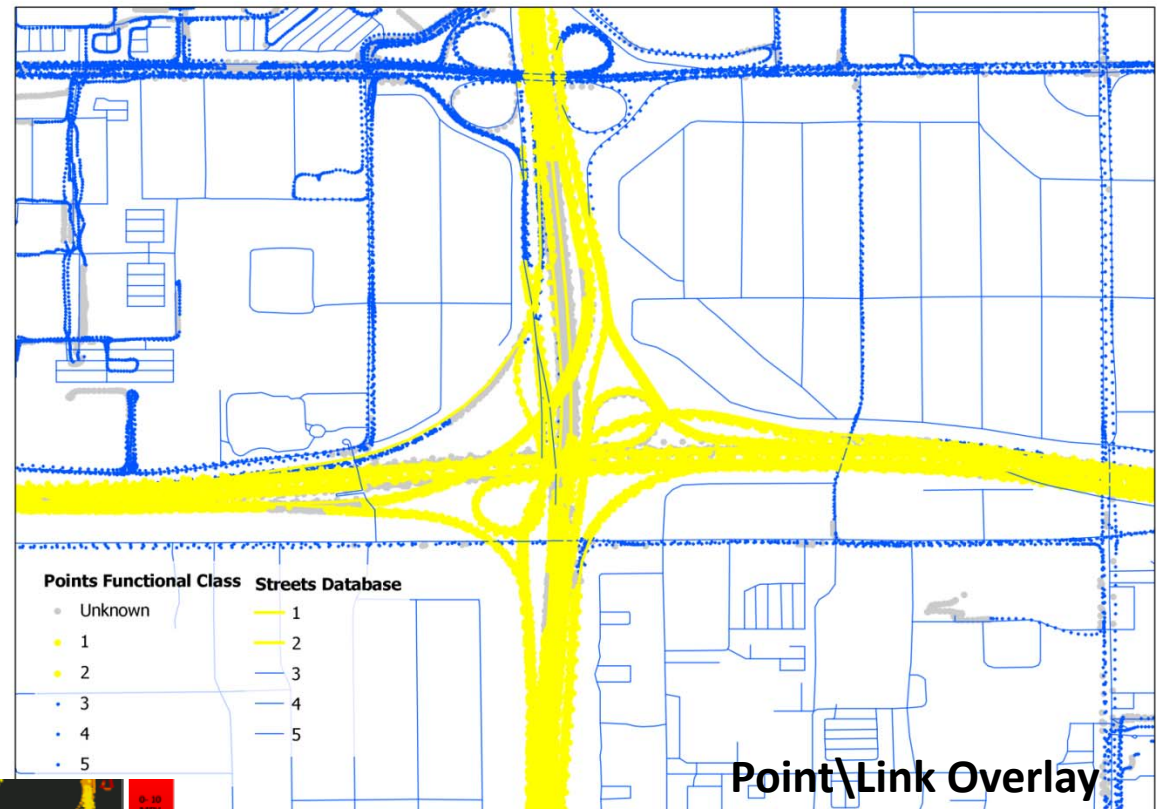
INTRODUCTION

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 Illustration

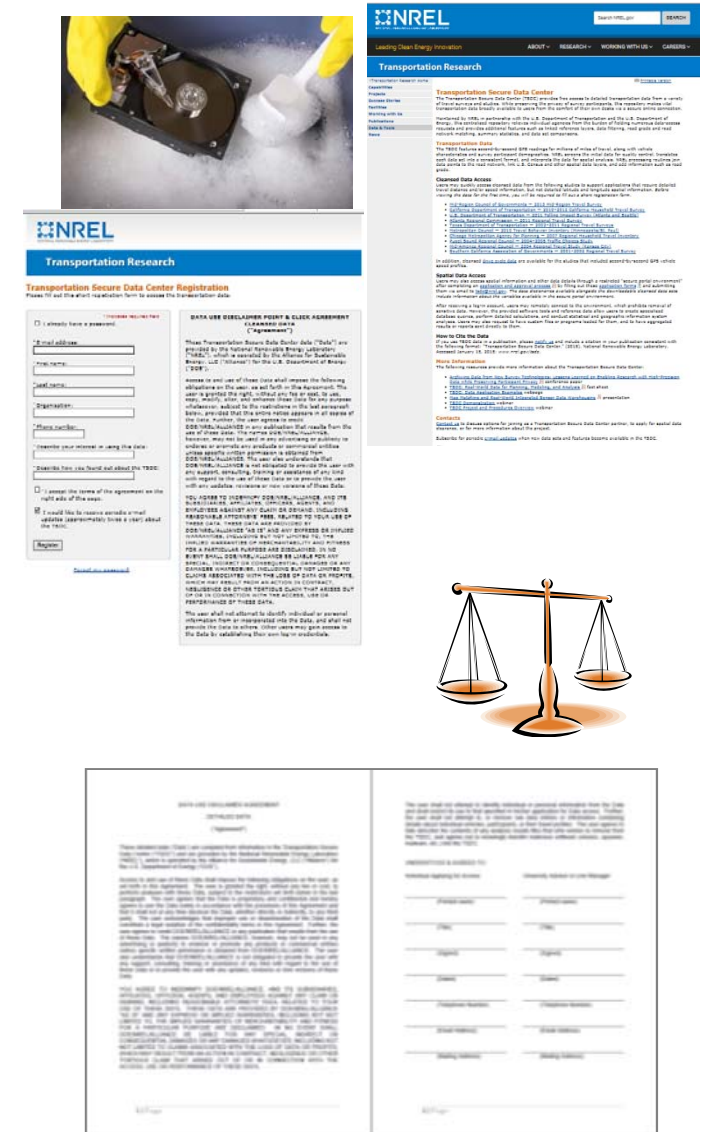
Complex overpasses

- *Connectivity can become ambiguous when so many options are available*
- *95% of distance matched across all data sets*
- *Cleaned up post processing during road based analysis*



TSDC Data Access: Established two distinct methods

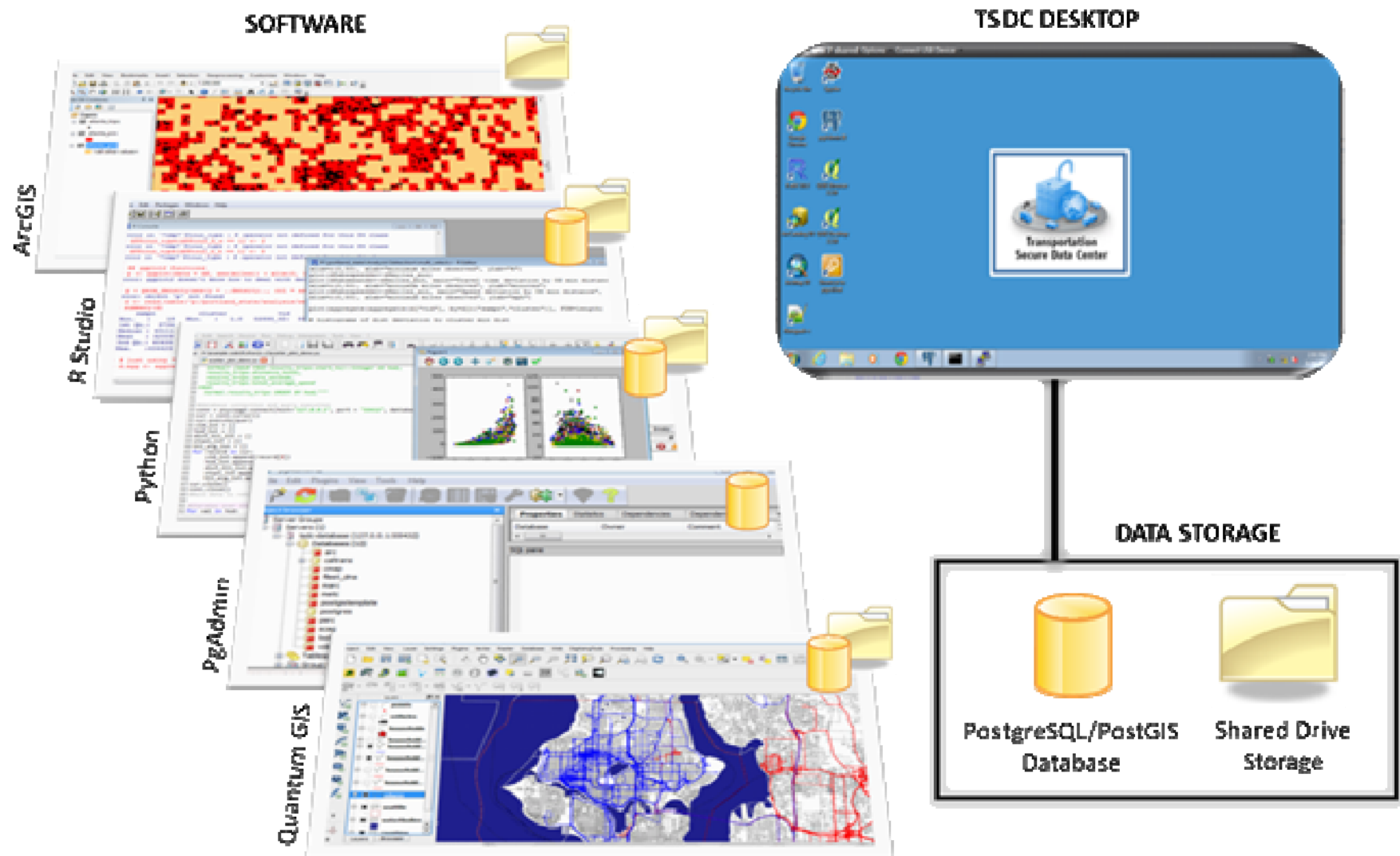
- Cleansed/public download data area
 - Streamlined access for cleansed data; helps limit accounts in secure portal to those with a legitimate need to work with the detailed data
 - Excludes latitude/longitude and other potentially identifying details (e.g., vehicle model)
 - Includes useful supplemental information (e.g., distance traveled by road type)
 - Requires point-and-click user registration and usage agreement
- Secure portal for detailed/spatial data
 - Virtual access (rather than requiring travel)
 - Details on next slide



Secure Portal Environment Access Process

- Application packet at www.nrel.gov/tsdc
- Data Use Disclaimer Agreement
 - Includes confidential data protection legal language and explicit pledge not to attempt identifying individual participants
 - Required for each individual user—no data removal or account sharing
 - Requires signature from both applicant and their supervisor
- Analysis Description Document
 - Explain proposed analysis, why secure portal access needed
- Condition of Use for NREL Cyber Resources (on-line form)
- Advisory group reviews application and provides recommendation
 - Data providers included on review if desired
- Approved users only access data within the secure portal environment
 - Data transfer prohibited (clipboard sharing, local drive access, & internet disabled)
 - Use software packages provided within the environment
 - NREL audits aggregated results a user wishes to remove before providing them to the user

TSDC Secure Portal Snapshot



Example Datasets

Example GPS Data Sets	# Vehicles	# Days	# Persons	# Days
2013 Mid-Region Council of Governments (Albuquerque) Travel Survey	NA	NA	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) Travel Behavior Inventory	NA	NA	174	7
2007 Chicago Regional Household Travel Inventory	408	7	209	7
2004–2006 Puget Sound Traffic Choices Study	484	540	NA	NA
2004 Mid-America Regional Council (Kansas City) Regional Travel Study	408	5	NA	NA
2001–2002 Los Angeles Regional Household Travel Survey	624	2	NA	NA
2002–2011 Texas Regional Household Travel Surveys	3,404	1	NA	NA

San Francisco, CA
Wearable GPS Trip Sample
By Mode of Travel

Color-coded routes illustrate travel mode trip segmentation.
Note: Does not include Bay Area Rapid Transit data.

- Caltrans data also includes OBD sample and geocoded trip ends from the full survey sample ($\approx 43K$ HH) in the secure portal environment

OBD = On-board diagnostic (information from the vehicle data bus including engine speed, etc.); HH = households

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

- If interested in partnering on the project
- For user support
- For help answering questions



**Transportation
Secure Data Center
(TSDC)**