

NSTSCCE

**National Surface Transportation
Safety Center for Excellence**

Linking Large Data Sets to Roadway Data

Shih-Ching Wu and Shane McLaughlin
Virginia Tech Transportation Institute

August 29, 2012



Outline



- Naturalistic Driving Study (NDS) Data
- Location-based Research Questions
- Linkage between NDS Data and Maps
 - ▣ Map Matching Algorithm
 - ▣ Large Scale Data Processing

Naturalistic Driving Study Data

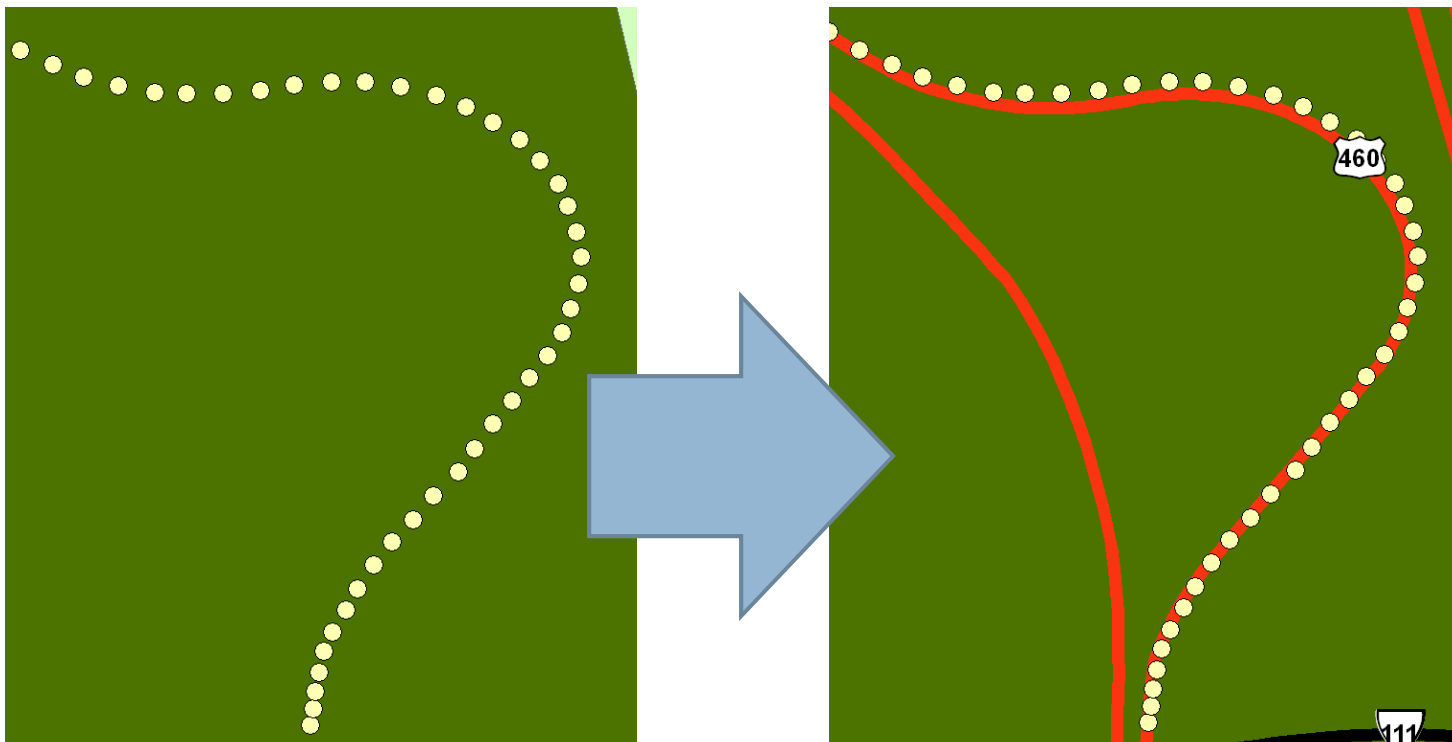
- Large amount of data
 - Terabyte ~ petabyte collections
- Variety of information collected from
 - Cameras, radar, vehicle network, GPS
- Do NOT collect location information directly from any devices
- Can be derived from GPS latitude and longitude coordinates

Location-based Research Questions

- Where did crash events occur the most?
- What's the percent of time or mileages traveled on each road type?
 - ▣ Interstate, U.S./primary/secondary highway, etc..
- Which areas/roads are people speeding?
- Intersection traverse
- Origin - destination

Map Matching Process

- Associating GPS traces on road segments



Knowledge of which road segments vehicles are on

NSTSCE

National Surface Transportation
Safety Center for Excellence

Elements of Map Matching Process



- Data
 - ▣ GPS data
 - ▣ Maps
- Method for conducting association
 - ▣ Map-matching algorithms
- Software
 - ▣ Geo-processing tools

Map Matching Cases

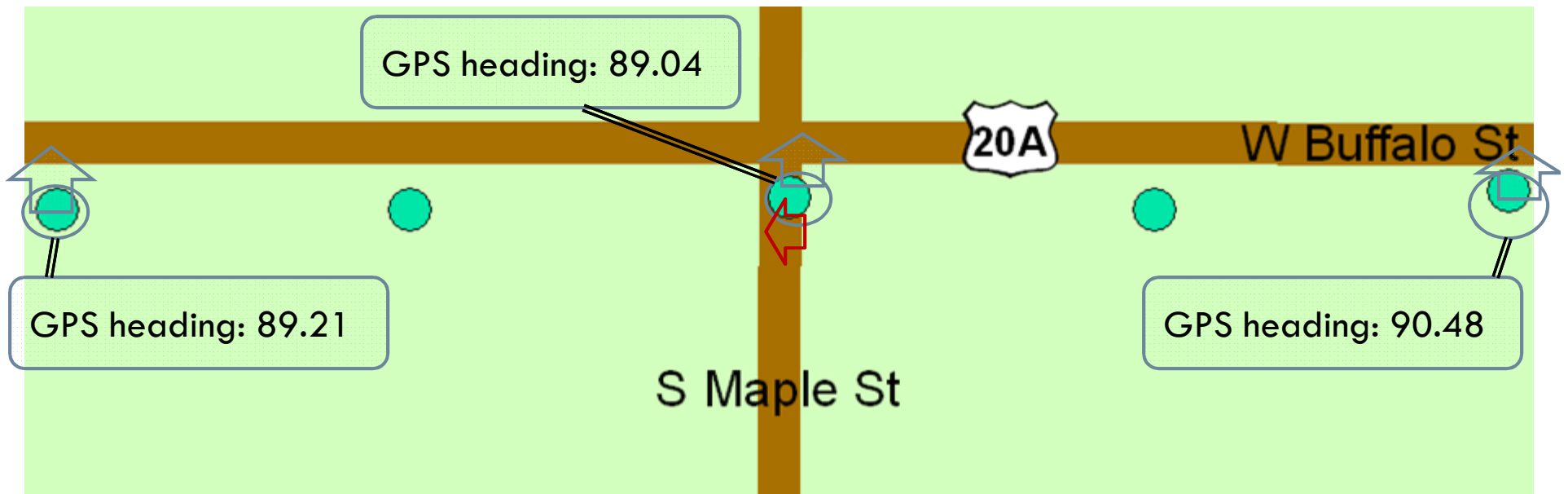
- Selection of candidate road segments based on spatial relationship
 - ▣ Shortest distance



Map source: Google maps

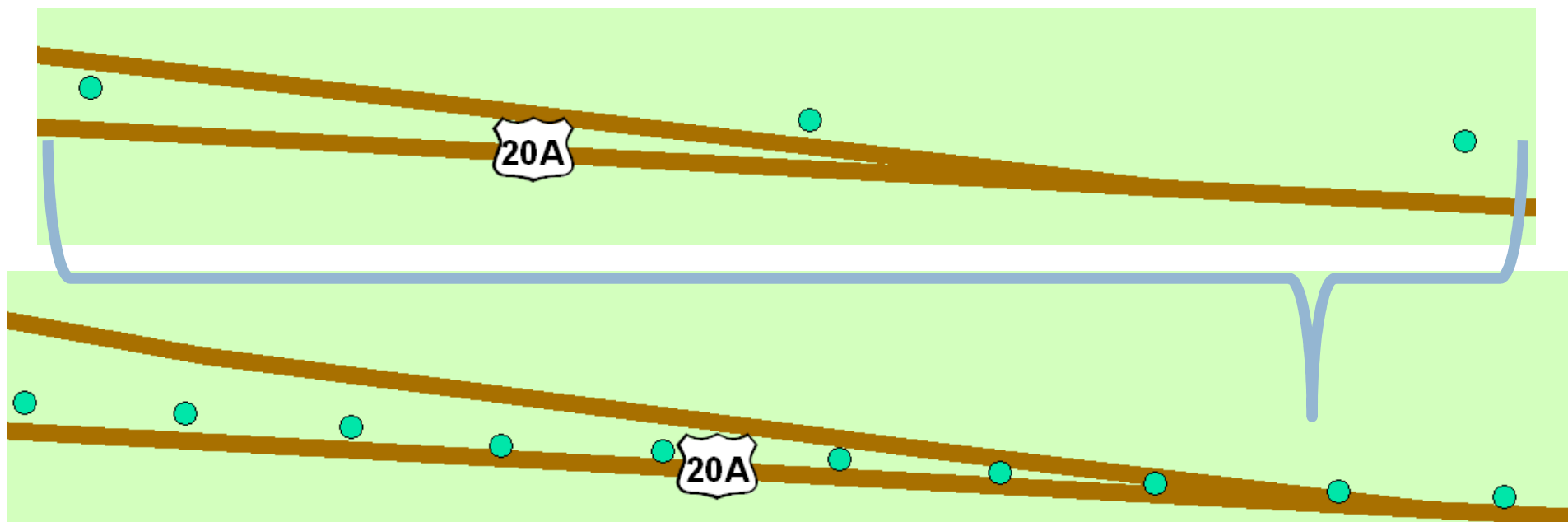
Map Matching Cases (Cont'd)

- Shortest distance
- Agreement among GPS heading to direction of roads



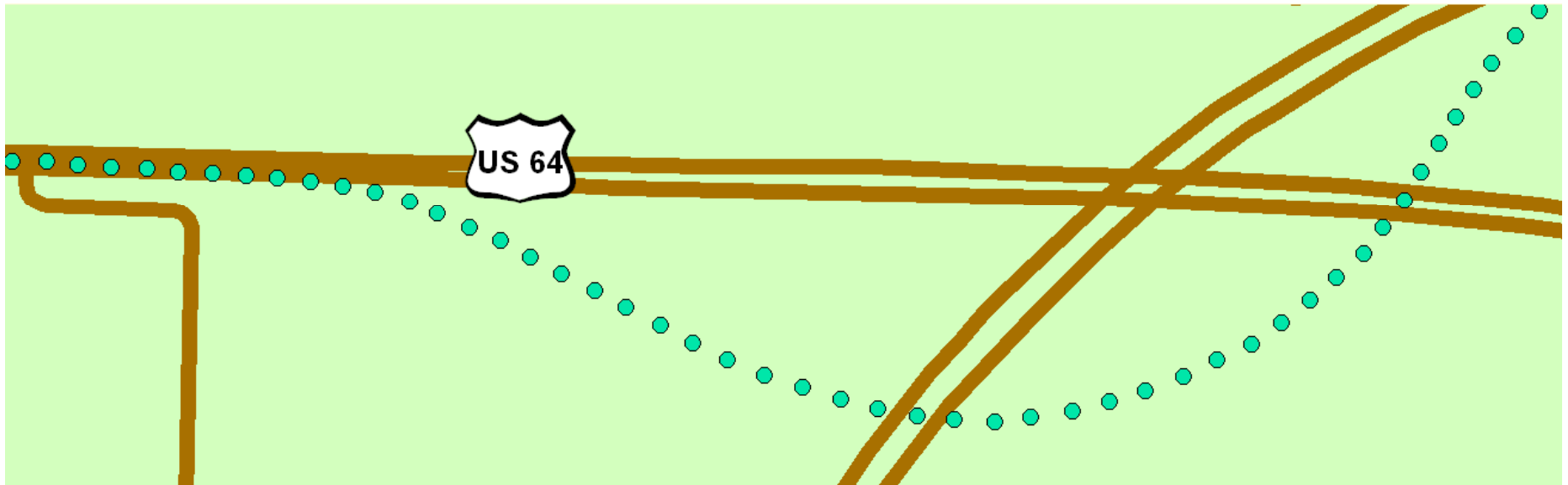
Map Matching Cases (Cont'd)

- Connectivity of map matching results



Map Matching Cases (Cont'd)

- Map resolution



Choices of Map Matching Parameters



- Distance between GPS and road segments
- Agreement among GPS heading and road directions
- Connectivity of map matching results
- Accurate map data

Implementation - Software



- GUI-based GIS software
 - ▣ Graphical user interface
 - ▣ Lack of flexibility to implement user-developed algorithms
 - ▣ Lack of capability to handle large amount of GPS data

Implementation - Software (Cont'd)

- Database application
- PostgreSQL
 - ▣ Database management systems
- PostGIS
 - ▣ Library of spatial database functions
- Simple scripting languages for constructing spatial queries
 - ▣ ST_Distance

Implementation - Software (Cont'd)

- Abilities to
 - ▣ implement user-developed algorithms
 - ▣ handle large-scale data
 - ▣ manage map matching processes

file_id	process_status	worker	file_process_date
8742624	2	LTSWU	2012-07-06
2121693	2	AWSMCLAUGHLIN	2011-12-29
3682956	0	AWSMCLAUGHLIN	2011-12-29
18723460	2	DT129CASR04	2012-07-06
3655875	2	TMVTTI003	2011-12-29
11442562	1	DT129CASR04	2012-07-06

Implementation - Software (Cont'd)

- Abilities to
 - ▣ implement user-developed algorithms
 - ▣ handle large-scale data
 - ▣ manage map matching processes

	file_id	timestamp	latitude	longitude	speed_mph	distance_feet	heading_gps	road_objectid	road_mapused	write_time
1	1892139	50975	42.████████	-77.████████	2.76163210	109.25403009	268.790008544922	0042	DOT	2012-03-30 15:30:26
2	1892139	72976	42.████████	-77.████████	11.92104522	44.96416183	274.970001220703	0042	DOT	2012-03-30 15:30:26
3	1892139	74976	42.████████	-77.████████	15.64924855	44.79189702	268.059997558594	0042	DOT	2012-03-30 15:30:26
4	1892139	76976	42.████████	-77.████████	19.80320377	57.07578619	269.730010986328	0042	DOT	2012-03-30 15:30:26
5	1892139	78976	42.████████	-77.████████	19.86073697	57.00786179	271.899993896484	0042	DOT	2012-03-30 15:30:26
6	1892139	79976	42.████████	-77.████████	19.61909496	32.84241468	271.209991455078	0042	DOT	2012-03-30 15:30:26
7	1892139	81976	42.████████	-77.████████	19.41197202	61.09569528	268.649993896484	0042	DOT	2012-03-30 15:30:26
8	1892139	83976	42.████████	-77.████████	21.57525156	59.30572639	260.140014648438	0042	DOT	2012-03-30 15:30:26

NDS data

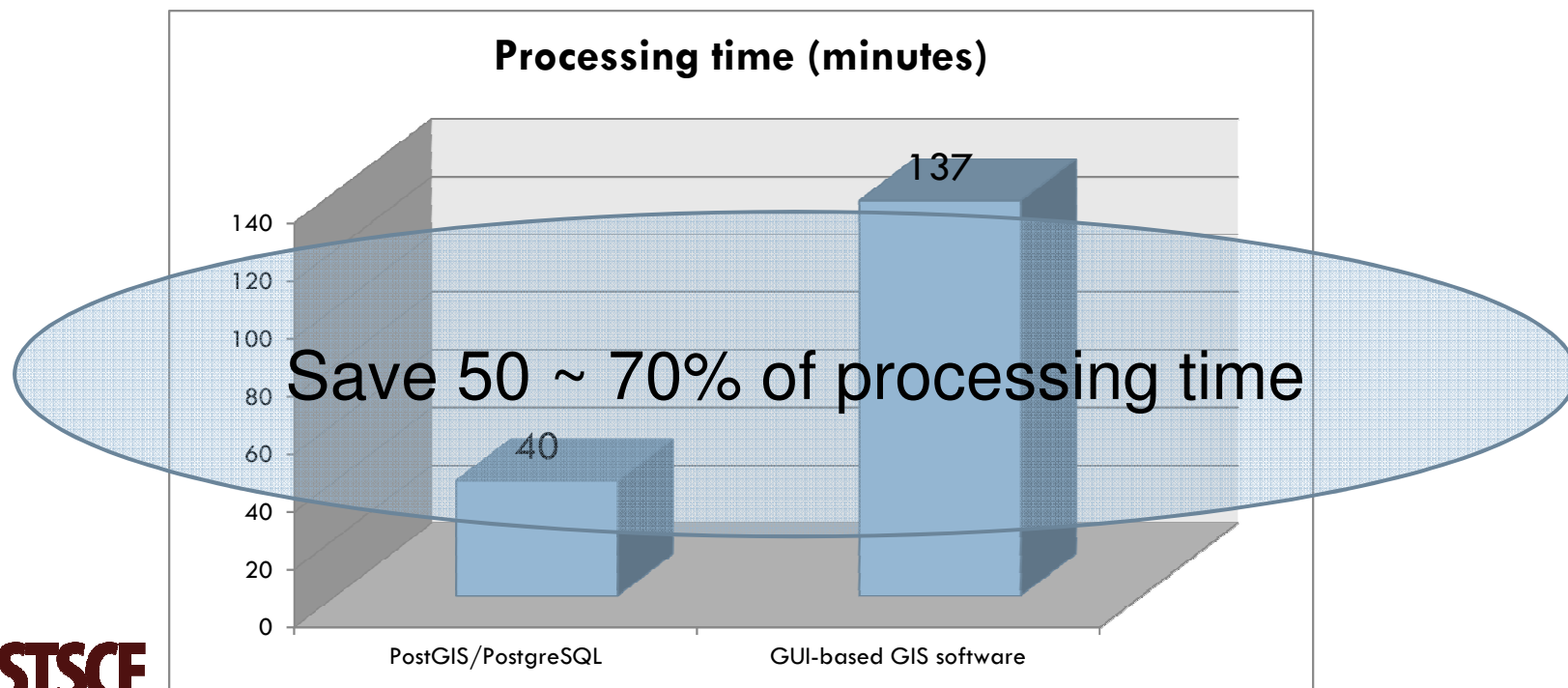
Roadway data

NSTSCE

National Surface Transportation
Safety Center for Excellence

Implementation - Software (Cont'd)

- Less processing time
 - ▣ Number of GPS points: 1,153,728
 - ▣ Number of road segments: 374,493 (NY)

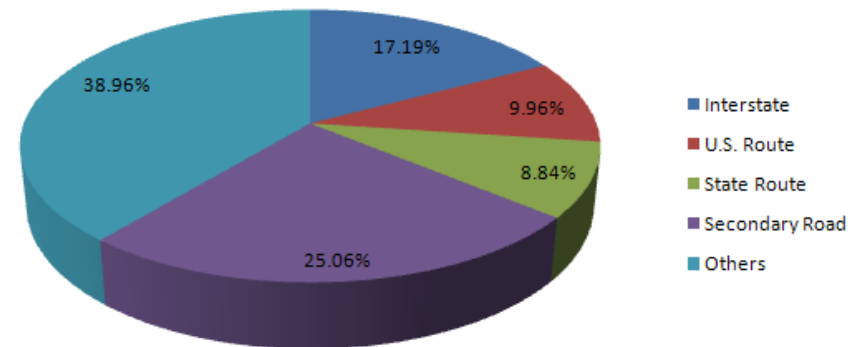




Questions?

Contact:
swu@vtti.vt.edu

Percent of Mileage on Each Road Class



References

- Shih-Ching Wu, Shane McLaughlin (2012) "Creating a Heatmap Visualization of 150 Million GPS Points on Roadway Maps via SAS", 20th Annual Southeast SAS Users Group Conference, Oct. 14-16, 2012
- Shih-Ching Wu, Shane McLaughlin (2012) "Tips for Using SAS to Manipulate Large-scale Data in Databases", 25th Annual Northeast SAS Users Group Conference, Nov. 11-14, 2012
- Yu Zheng, Xiaofang Zhou (2011) "Computing with Spatial Trajectories", 1st Edition, Springer

References (Cont'd)

- Strategic Highway Research Program 2
 - ▣ <http://www.shrp2nds.us/>
- PostgreSQL
 - ▣ <http://www.postgresql.org/>
- PostGIS
 - ▣ <http://postgis.refrations.net>