RELATING NATURALISTIC GLOBAL POSITIONING SYSTEM (GPS) DRIVING DATA WITH LONG-TERM SAFETY PERFORMANCE OF ROADWAYS

Western States Forum 2017

Anurag Pande, Ph.D.
Associate Professor
Cal Poly San Luis Obispo

Steve Lakowske
Vice President, Telematics
eDriving

Wednesday, July 26, 2017

About Transportation @Cal Poly

- Best ITE (Institute of Transportation Engineers) Student Chapter in the Western district four years in a row

- Part of the CTEDD (Center for Transportation Equity Decisions and Dollars) University Transportation Center Consortium (2017-2022) that provides federal matching funds for applied research

- Faculty expertise in Measurement of Safety, Resilience, Traffic Engineering, and Transportation Economics

Overview

- Introduction & Background
- Naturalistic Driving Data Collection & Processing
- Identification of Crash Prone Highway Segments
- Considerations for scaling the effort: eDriving
- Questions & Answers

Introduction

- Traffic safety analysis relies on historical traffic crashes to identify hazardous roadway locations
  - Collected over long periods of time (5 to 10 years)

- Recent research efforts seek to identify better and more efficient methods to identify hazardous roadway conditions
  - Use of naturalistic driving studies
  - Identification of near crashes or crash surrogates

Background Studies

- 100-Car Study
  - Performed by Virginia Tech Transportation Institute and NHTSA
  - Extensive instrumentation on board vehicles (Dingus et al., 2005)
  - Studies involving risks associated with driver inattentiveness (Klauer et al., 2006)

- Other research utilizing rate of change of acceleration and sudden braking to identify crash surrogates (Bagdadi, 2013)

- Current Research – SHRP2

Study Objectives

- Research study funded by the National Science Foundation (NSF) to utilize naturalistic GPS driving data
  - Establish a methodology for processing large GPS driving data sets collected from a naturalistic driving study
  - Identify measures that correlate with long-term safety performance of highways
  - Calculate these measures from the GPS data and explore statistical links between driving measures and long-term traffic safety performance

Study Differences

- The data used for this analysis is obtained from a simple GPS device and not from extensive instrumentation (e.g., 100-Car Study)
- Driving data is being related to long-term safety performance and is not used for the real-time identification of near-crashes or crash surrogates

Data Collection

- GPS Devices - OHARA Corp. GPS Data Loggers V3.15

Data Collection

- The total data collection period lasted from July 2012 to March 2013.
- Each participant in the study was given a GPS device for their personal commute vehicle for a period of approximately two weeks.
- To preserve battery, each data logger was programmed with a sleep mode that disabled data recording if the device remained idle for a period greater than 300 seconds.
- Typical battery life of each GPS data logger was 9 to 11 complete days before recharge is required.
- Latitude and longitude were recorded by the devices to the standards of the 1984 update of the World Geodetic System (WGS84).

Device Details

- The devices were powered by a rechargeable 1800 mAh 3.6V lithium ion battery.
- An 8 gigabyte micro SD memory card stored the collected data while a multi-color LED light indicated functionality of the data logger.
- The circuit board and GPS chip were protected in a protective case to assure they were not damaged during the data collection.

GPS data attributes

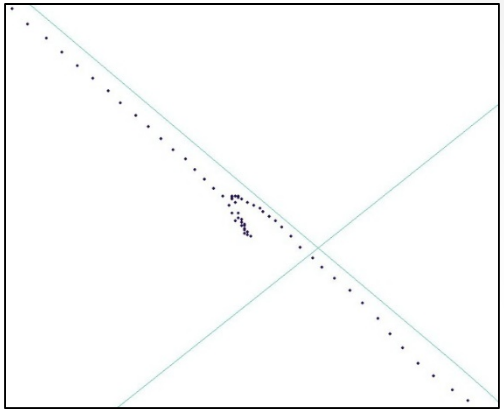
Attribute	Attribute Description	Remark / Unit
LATITUDE	Latitude	WGS84, Degrees (°)
LONGITUDE	Longitude	WGS84, Degrees (°)
ALTITUDE	Altitude	WGS84, Meters (m)
HEADING	Directional Heading of Movement	Degrees (°)
SPEED	Velocity of Device	Miles Per Hour (MPH)
SAT	Number of Satellites in Communication	
PDOP	Positional Dilution of Precision	
HDOP	Horizontal Dilution of Precision	
VDOP	Vertical Dilution of Precision	
FIX	Recording Status	

GPS device placement

- Typically, the GPS loggers were positioned in vehicle center consoles or glove boxes.
- Through initial testing it was determined that placing the logger in either location did not impact GPS communication.
- Values of HDOP and PDOP remained below a value of 2 for over 85% of the data from all participants.

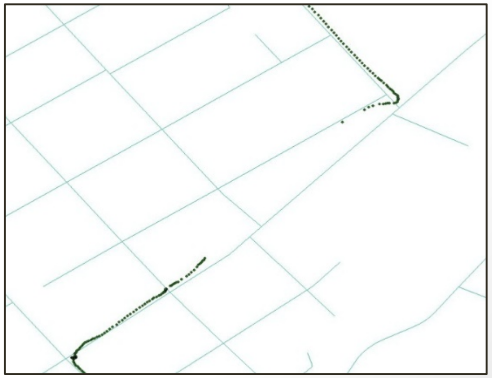
Data Errors

- Noise,

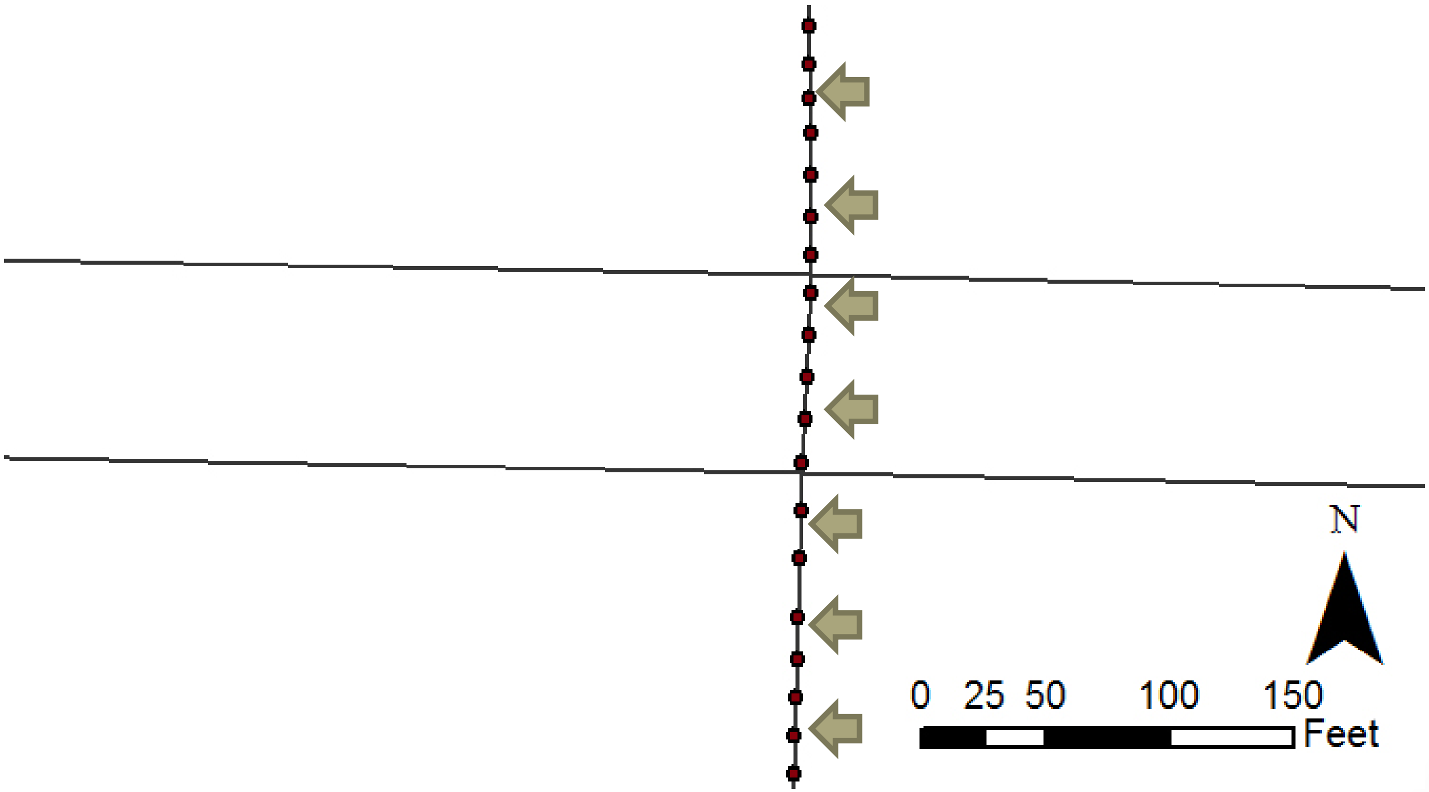


- Wandering,

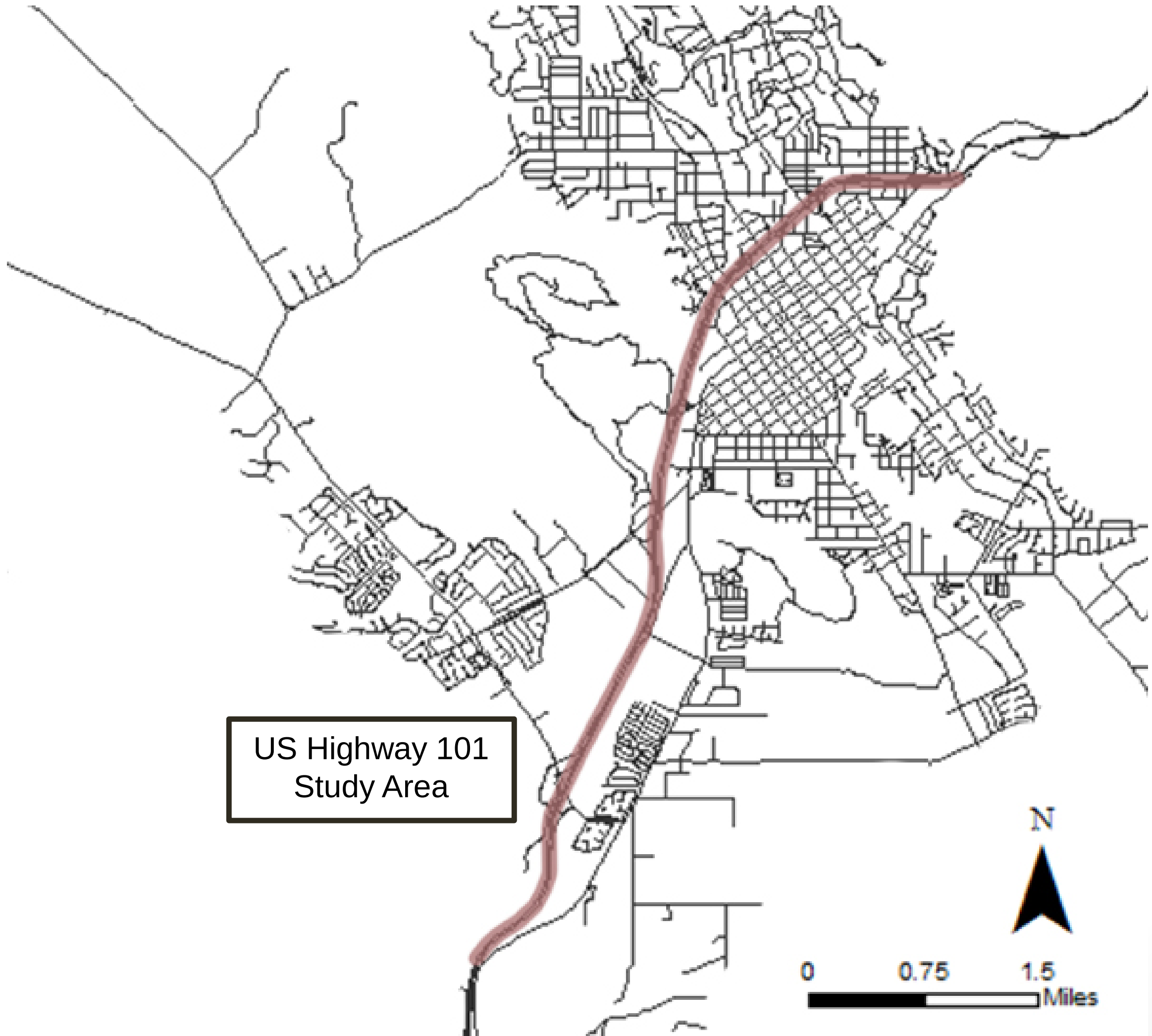
- Gaps.



Data Processing using GIS – Linear Referencing



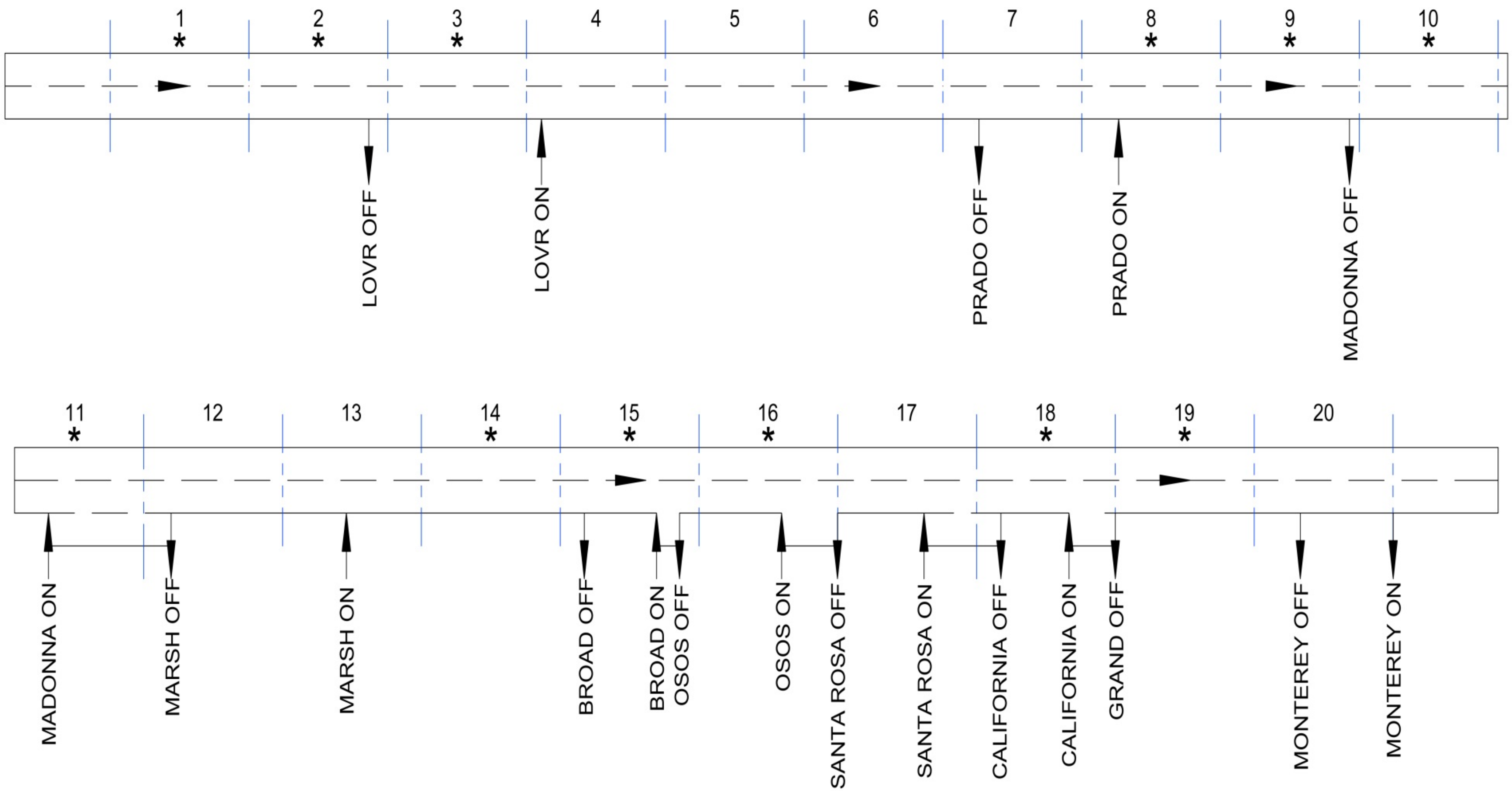
Study Area: US 101 freeway in San Luis Obispo



Highway Analysis

- Vehicle crash data collected from Transportation Injury Mapping System (TIMS) network maintained by UC Berkeley
- Collected all crashes along US Highway 101 from 2002 to 2011
- Total number of crashes converted into a measure of crash rate by taking into account ADT for each section based on 2011 traffic volumes
- Analysis performed for 39 quarter-mile segments and 19 half-mile segments of the US Highway 101

US 101 NB ½-mile Segments



Estimation of Jerk

- Calculation of Acceleration and Jerk:

$$a = \frac{\Delta v}{\Delta t} \qquad\qquad j = \frac{\Delta a}{\Delta t}$$

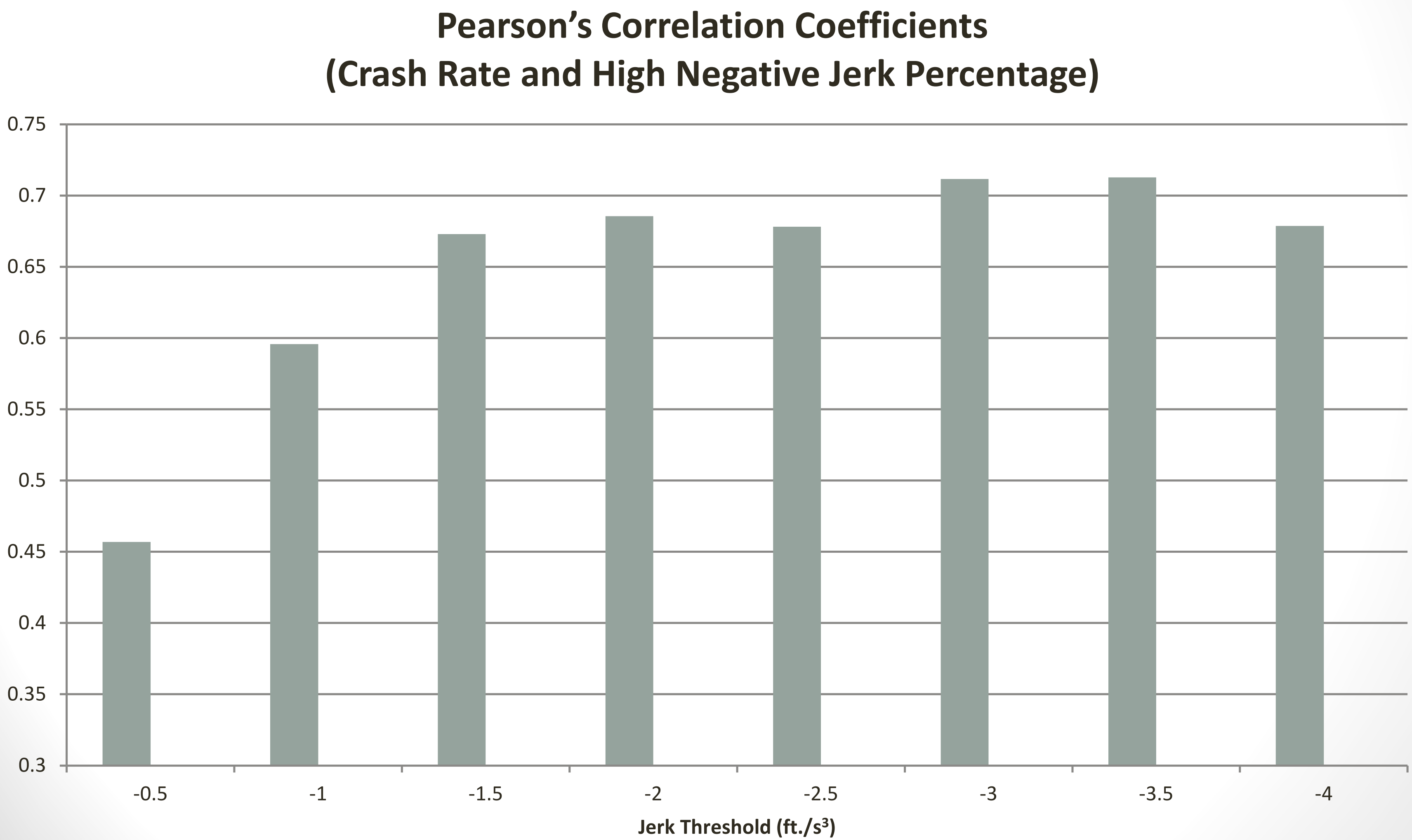
where:

a:	Acceleration (ft/s^2)	Δv:	Change in velocity (ft/s)
Δt:	Change in time (s)	j:	Jerk (ft/s^3)
Δa:	Change in acceleration (ft/s^2)		

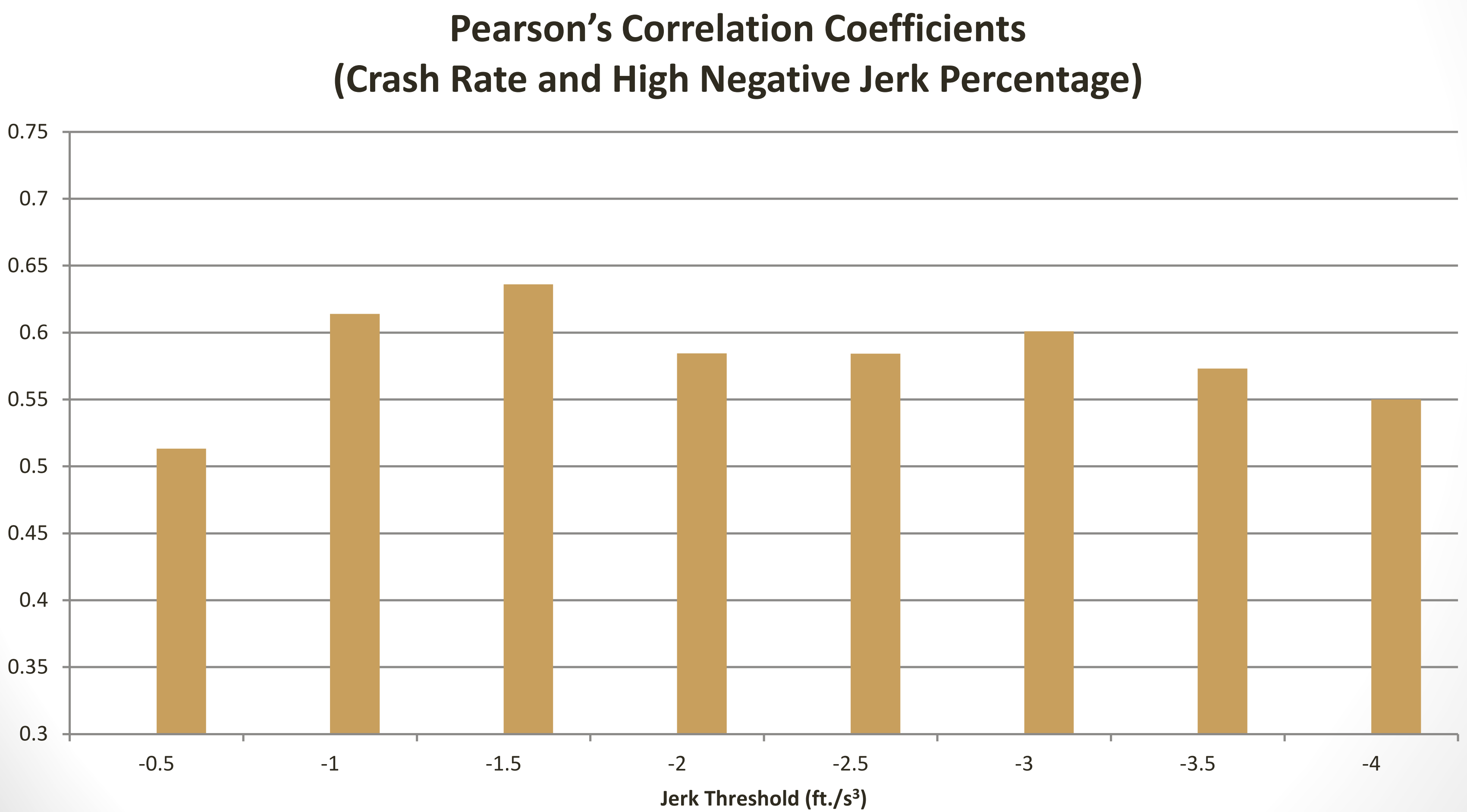
Jerk Estimate Accuracy

- Accuracy of the jerk measure derived from GPS data
- The analysis does not use Jerk values directly into the model just percentage higher than a threshold
- Interested in percentage of observations with jerk value greater than a certain threshold level for each segment
- Following slides show the correlation between this percentage (@varying thresholds) and long-term crash rates

¼-mile US 101 Segments



½-mile US 101 Segments



Statistical Analysis

- The next question: Can the jerk percentage on a segment from this data explain long-term crash frequency of that segment
  - Estimate Negative binomial regression model for crash frequency on ¼-mile segments
  - Attempt to explain crash frequency (the dependent variable) with the following variables:
    - ADT of the segment
    - Curvature
    - Presence of Auxiliary lane

Crash Freq. Regression model for ¼-mile US 101 Segments

Negative Binomial Model with ADT and Geometric Variables Analysis of Maximum Likelihood Parameter Estimates			
Parameter	Estimate	Standard Error	p-value
Curve Presence	0.0183	0.2532	0.9418
Auxiliary Lane Presence	0.3008	0.3236	0.3526
Average Daily Traffic	1.6972	2.7316	0.5344

High p-values (>0.10) indicate that the coefficient for the variable in the model for crash frequency is not significantly different than zero. None of the variables are significant in the model shown above.

Crash Freq. Regression model for ¼-mile US 101 Segments

Negative Binomial Model with High Jerk Percentage Analysis of Maximum Likelihood Parameter Estimates			
Parameter	Estimate	Standard Error	p-value
Percentage of Observations with Jerk < -2 ft./s^3	0.1297	0.0449	0.0038

Low p-value (<0.10) indicate that the coefficient for the variable in the model for crash frequency is significantly different than zero. The results are similar for all thresholds above -1.5 ft./s^3

Conclusions

- Parameters of interest could be effectively estimated from naturalistic GPS driving data using GIS linear referencing and data filtering
- The proportion of observations with high negative jerk percentage on highway segments was correlated with long-term crash rate on quarter-mile and half-mile segments
- In fact, the analysis showed that other measures (such as presence of curves, presence of auxiliary lane, and ADT) were insignificant or not as reliable for estimating long-term crash frequency

How do we scale this now?

- Methodology developed here can be suitable with multiple data sources for long-term safety assessment
  - SHRP2 naturalistic driving data
  - Data from cellular devices
  - Commercial fleet GPS data

“Commodity” Data Collectors

- Mobile (CMT, ZenDrive, Mentor, TrueMotion...)
  - Pros: Inexpensive, pervasive, multipurpose, flexible/programmable, full data set can be recorded, any format
  - Cons: Orientation calc required if not cradled/fixed, non-deterministic operating system, varied sensor quality, no OBD unless hybrid config.
- OBDII (GeoTab, Danlaw, Davis, CalAmp…)
  - Pros: High frequency sampling, accurate, OBD parameters, other sensors
  - Cons: Cost, typically event-triggered recording, data plan usually needed
- Black Box (CalAmp, Octo, Custom…)
  - Pros: Highest accuracy/precision/sampling rate, tamper-resistant, additional interfaces
  - Cons: Cost, installation, access
- Hybrid (VTTI…)

Smartphones as Data Collectors

- Benefits
  - Comprehensive Sensor Suite (accelerometer, gyro, gps)
    - Compare different sensor data quality
    - Redundancy, Sensor Fusion
  - Interfaces Built in (BT/BLE, WiFi, Cellular)
    - Connect to Vehicle OBDII Bus via Scanner
    - Transmit Data to Cloud/Server
- Caution
  - Fixed Orientation for Best Results, w/Auto-Calibration Routines
- Hybrid Mobile-OBD (scanner) Best Price/Performance
  - 10 hz OBD + smartphone sensors

Technical Considerations

- GPS
  - Asynchronous data sampling (delta-t varies)
  - Smartphone best rate usually 1/sec max
    - 10/sec possible w/external GPS
- Accelerometers, Gyros, Magnetometer
  - Magnetometer impacted by mag field disruption in vehicles
  - Smartphone OS not deterministic
  - High-frequency sampling & filtering/derivation on device
  - Select and use filters correctly
  - Data volume/sample rate/precision tradeoff

edriving

Telematics Data Spectrum

continuous time-series data

Tier 1

- timestamp
- latitude
- longitude
- gps speed
- pdop
- satcnt
- grav-x
- grav-y
- grav-z
- gyro-pitch
- gyro-roll
- phone in/out
- text in/out
- obd_speed
- obd_throt
- obd_rpm
- obd_maf
- accelx
- accely
- accelz

Hybrid

Tier 2

- timestamp
- latitude
- longitude
- gps speed
- pdop
- satcnt
- etc.
- grav-x
- grav-y
- grav-z
- gyro-pitch
- gyro-roll
- phone in/out
- text in/out

Tier 3

- timestamp
- latitude
- longitude
- gps speed
- obd_speed
- events
- pdop
- satcnt

OBDII (min)

Tier 4

- timestamp
- latitude
- longitude
- gps speed

Minimum

summary data

Tier 5

- timestamp
- A events
- B events
- C events
- D events
- S events
- tripid
- driverid
- duration
- distance
- speedmax
- etc.

Tier 6

- timestamp
- A events
- B events
- C events
- S events
- driverid
- duration
- distance
- etc.

TSP

Device ID
Group
of Idling Events > 5 mins
Idling Duration
% Idle Time
Tamper Signs
Total Distance(Miles)
Total Drive Time
Total Stops
Speeding Events > 10% Over Limit
Speeding Events > 15% Over Limit
Speeding Events > 20% Over Limit
SA #
HB #
Speed #
of Aggressive Events
Aggressive Events/100 Miles

variables for ABCS calc
variables for distraction calc
variables for accurate AB accels-OBDII scanner w/accelerometer & TSP events

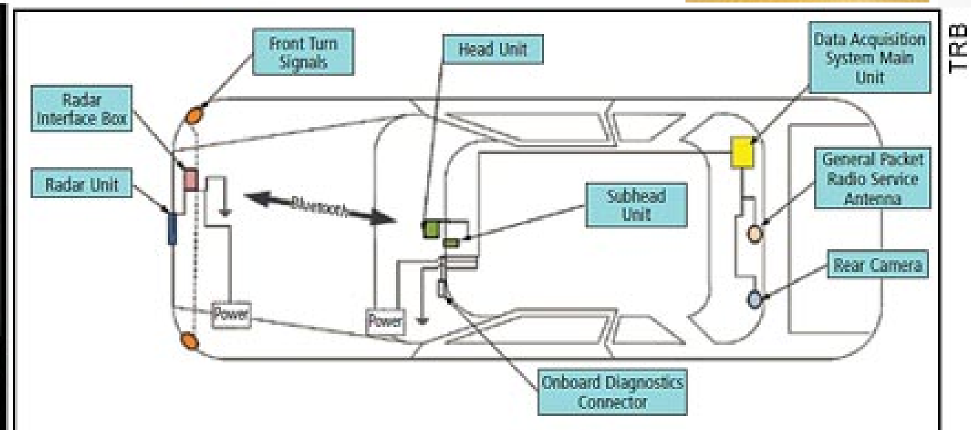
edriving

Industry Needs

- Standards
  - Parameters
  - Precision/Resolution
  - Accuracy
  - Frequency
- Data Quality
  - Objective Measures
  - Standard V&V Process

SHRP2 NDS Data Collector

SHRP2 Dataset



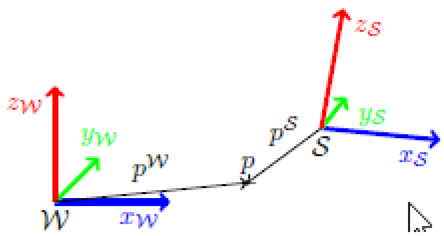
Data Acquisition System Channels

- Multiple videos
- Machine vision
  - Eyes forward monitor
  - Lane tracker
- Accelerometer data (3 axis)
- Rate sensors (3 axis)
- GPS: latitude, longitude, elevation, time, velocity
- Forward radar
  - X and Y positions
  - X and Y velocities
- Cell phone
  - Automatic collision notification, health checks, location notification
  - Health checks, remote upgrades
- Illuminance sensor
- Infrared illumination
- Passive alcohol sensor
- Incident push button—audio (only on incident push button)
- Turn signals
- Vehicle network data
  - Accelerator
  - Brake pedal activation
  - Automatic braking system
  - Gear position
  - Steering wheel angle
  - Speed
  - Horn
  - Seat belt information
  - Airbag deployment
  - Many more variables

edriving

References

Orientation Estimation Using Smartphone Sensors
https://www.control.isy.liu.se/student/tsrt14/file/orientation.pdf



Teaching Sensor Fusion and Kalman Filtering using a Smartphone
http://liu.diva-portal.org/smash/get/diva2:741780/FULLTEXT01.pdf

Driving Style Recognition Using a Smartphone as a Sensor Platform
https://cvrr.ucsd.edu/publications/2011/Johnson_ITSC2011.pdf

VTTI SHRP2 NDS InSight Web Site
https://insight.shrp2nds.us/

Acknowledgements

- James Loy, Cal Poly
- Russell White, Cal Poly
- Sean Carney & Nathan Johnston, Cal Poly
- All of our Volunteer Drivers

- Dr. Vinayak Dixit, University of New South Wales
- Dr. Brian Wolshon, Louisiana State University
- Joshua Kent, Louisiana State University

- National Science Foundation

CAL POLY
SAN LUIS OBISPO

LSU
LOUISIANA STATE UNIVERSITY

NSF

Questions?

Stephen Lakowske

VP Telematics

eDriving

Cal Poly Technology Park

San Luis Obispo, CA 93407

Phone: (303) 204-0192

E-mail: *steve.lakowske@edriving.com*

Anurag Pande

CEENVE

1 Grand Ave

San Luis Obispo, CA 93407

Phone: (805) 756-2104

E-mail: *apande@calpoly.edu*