Using Transportation Big Data in Safety Analytics

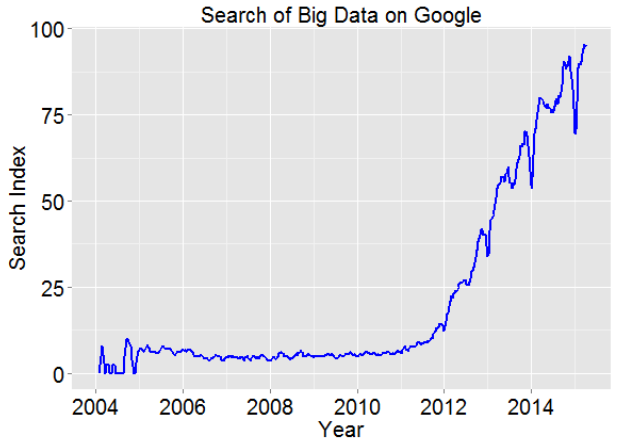
Dr. Mohamed Abdel-Aty, P.E.
Pegasus Professor and Chair
Department of Civil, Environmental & Construction Engineering

Deputy Director
Center for Advanced Transportation Systems Simulation
University of Central Florida, USA

Editor-in-Chief
Accident Analysis & Prevention

The Age of Big Data

- Age of Big Data
  - Since 2010
- Key words related to Big Data
  - Data
  - Big
  - Information
  - Predictions



*Data Source from Google Trends

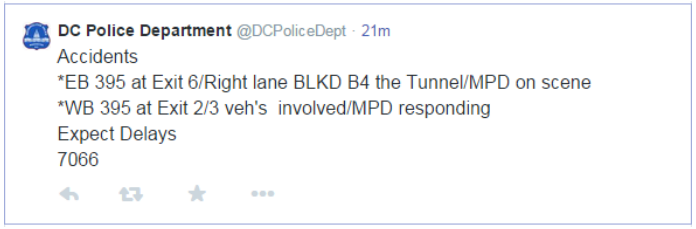
*Data Source from Big Data: A revolution That Will Transform How We Live, Work, and Think

What is Big Data?

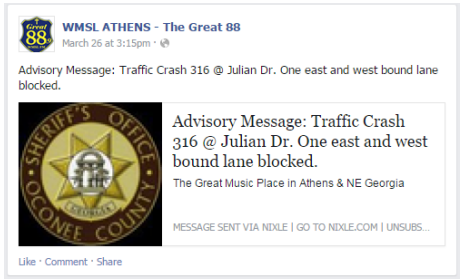
- General impression: storage and analysis of unfathomable amounts of information
- The most widely cited definition: three Vs (basic)
  - Volume: increasing size of data
  - Velocity: unprecedented streaming speed of data
  - Variety: wide range of data formats
- Additional dimensions of Big Data (sometimes used)
  - Veracity: trust and uncertainty with regards to data and the outcome of analysis of that data
  - Variability: data flow inconsistency with periodic peaks
  - Complexity: link, match, cleanse and transform data across systems

Big Data in Transportation Arena

- Sources of Big Data in transportation
  - Web traffic, network comments
  - ITS facilities (traffic sensors)
  - Roadway geometric database
  - Crash database
  - GIS database
- Data type
  - Structured data
    - Traffic detection data
    - Road geometric data
    - Crash data
    - Land use data
  - Unstructured data
    - Video images
    - Social media text
    - Text documents



*Data source from Twitter



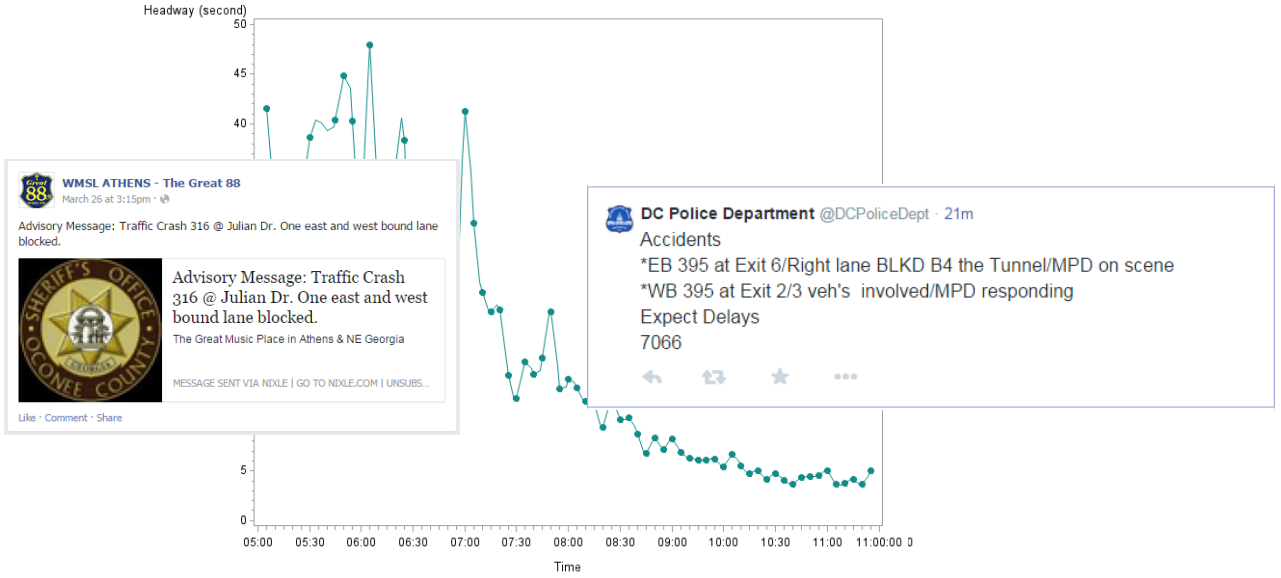
*Data source from Facebook

*Automatic Vehicle Identification (AVI) System (above)

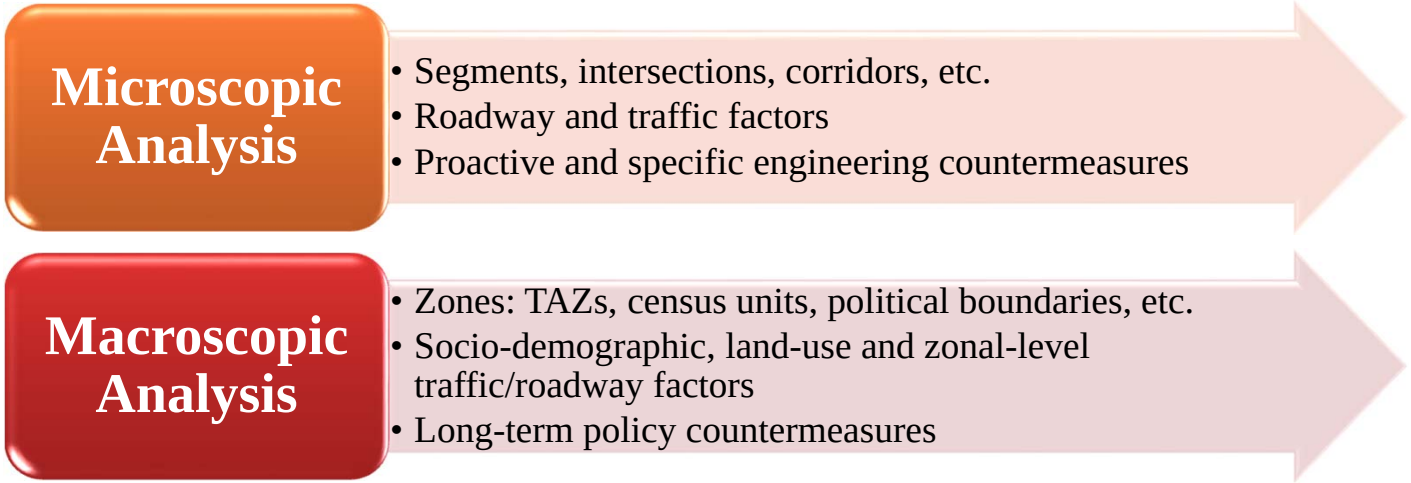
*Microwave Vehicle Detection System (MVDS) (right)

Features of Traffic Big Data

- Features of traffic Big Data
  - Data are often available in real-time
  - Data are available on previously unmeasured activities
  - Data come with less structure



Safety Analytics using Big Data



- Data mining solutions for Big Data safety analytics
  - Advantages: high predictive fit, especially out-of-sample fit
  - Common data mining techniques
    - Classification and regression trees
    - Random forest
    - Neural network
    - Support vector machine
    - Multivariate adaptive regression splines Clustering
    - K-nearest neighbors clustering

Safety Analytics using Big Data

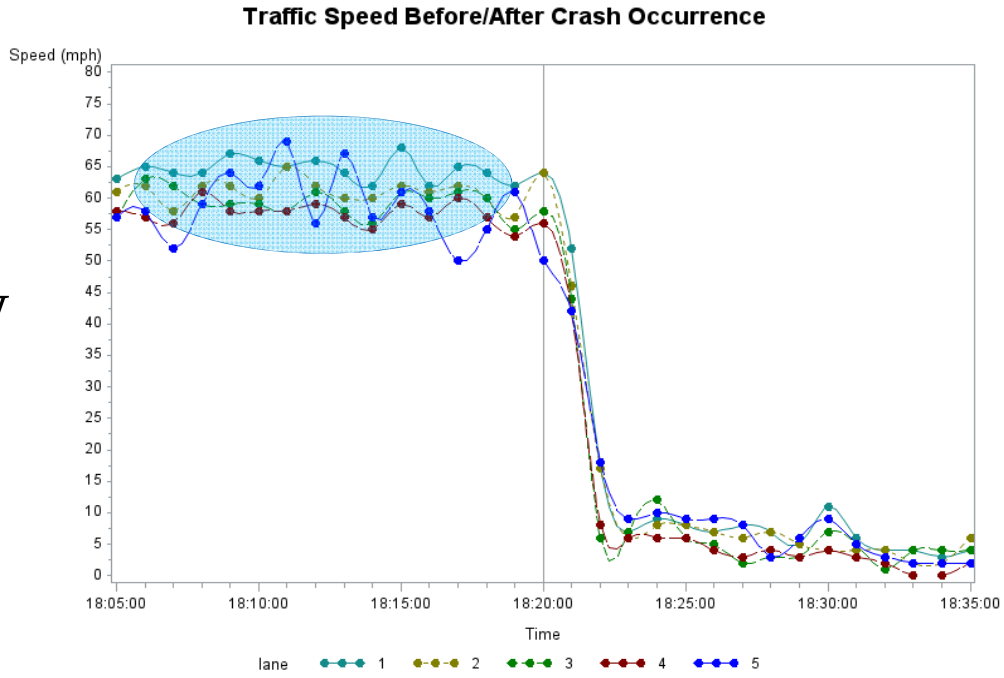
- Classification and regression trees
  - Machine learning optimal discriminant analysis
  - Classification of safe/unsafe traffic conditions, crash types and severity
- Random forest
  - Constructing a multitude of decision trees to avoid overfitting
  - Able to deal with large number of input variables without variable deletion
  - Estimating variable importance in the classification
- Neural networks
  - Nonlinear sophisticated modeling for complex problems
  - Useful when the exact nature of the relationships between inputs and output is unknown

Safety Analytics using Big Data

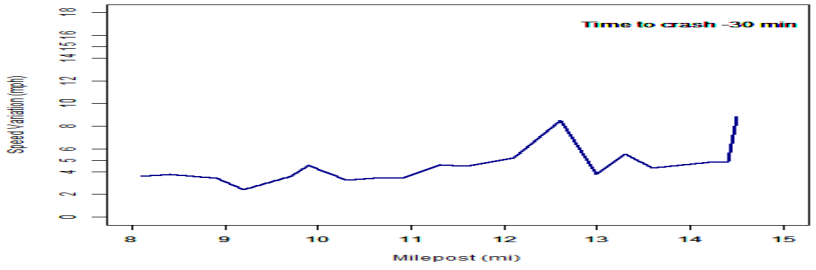
- Support vector machine
  - Detect and exploit complex patterns in data
  - Methods: clustering, classifying and ranking data
  - Useful for text and hypertext categorization (for social media data)
- Multivariate adaptive regression splines (MARS)
  - A non-parametric regression technique
  - Flexible to model non-linearity and variable interactions
  - Able to handle large datasets
- K-nearest neighbors (KNN) Clustering
  - Determine the class membership of observations that share higher similarities
  - Useful for generalizing traffic/weather conditions, and social media information for detailed safety analysis

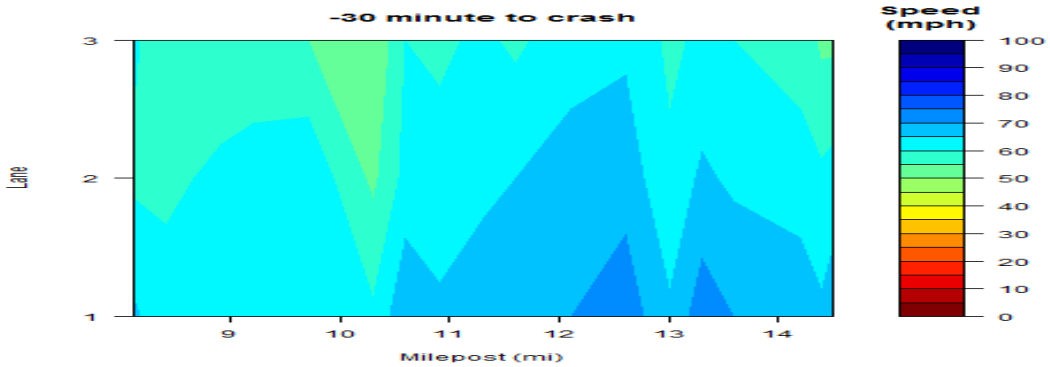
Big Data Applications in Microscopic Traffic Safety Analysis

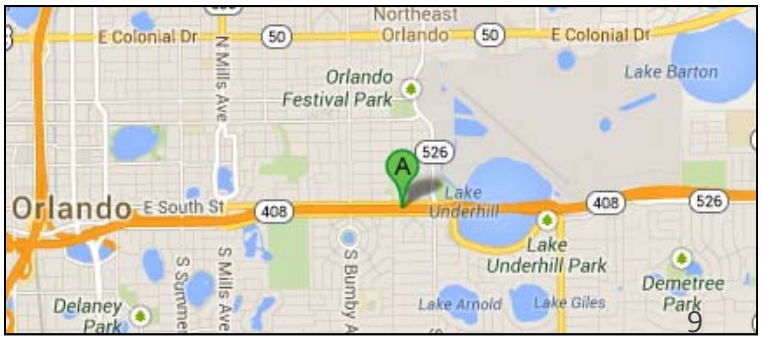
- Real-time nature of Big Data to reflect the occurrence of crash and their effects on traffic flow



Direction: WB, @MM:12.1, Type: Rear-end, Vehicles: 4, Injuries: 3

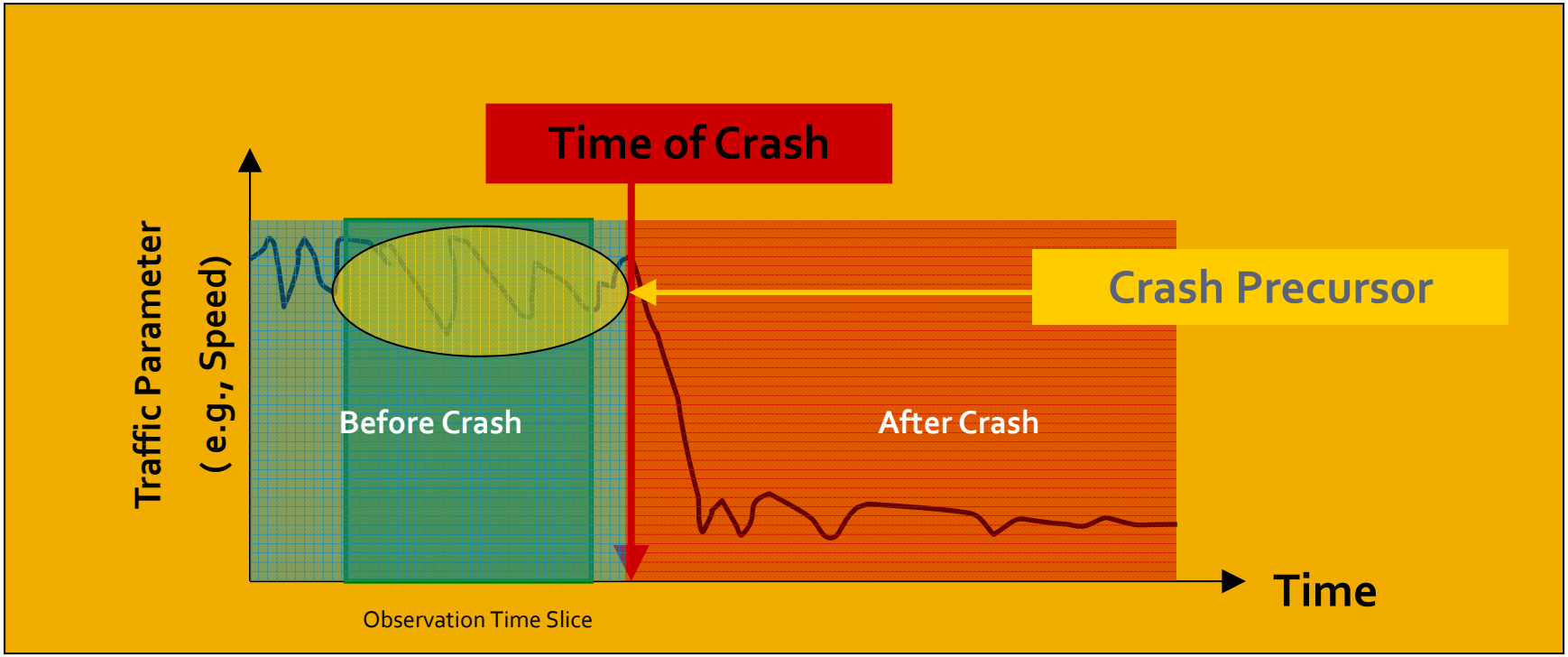






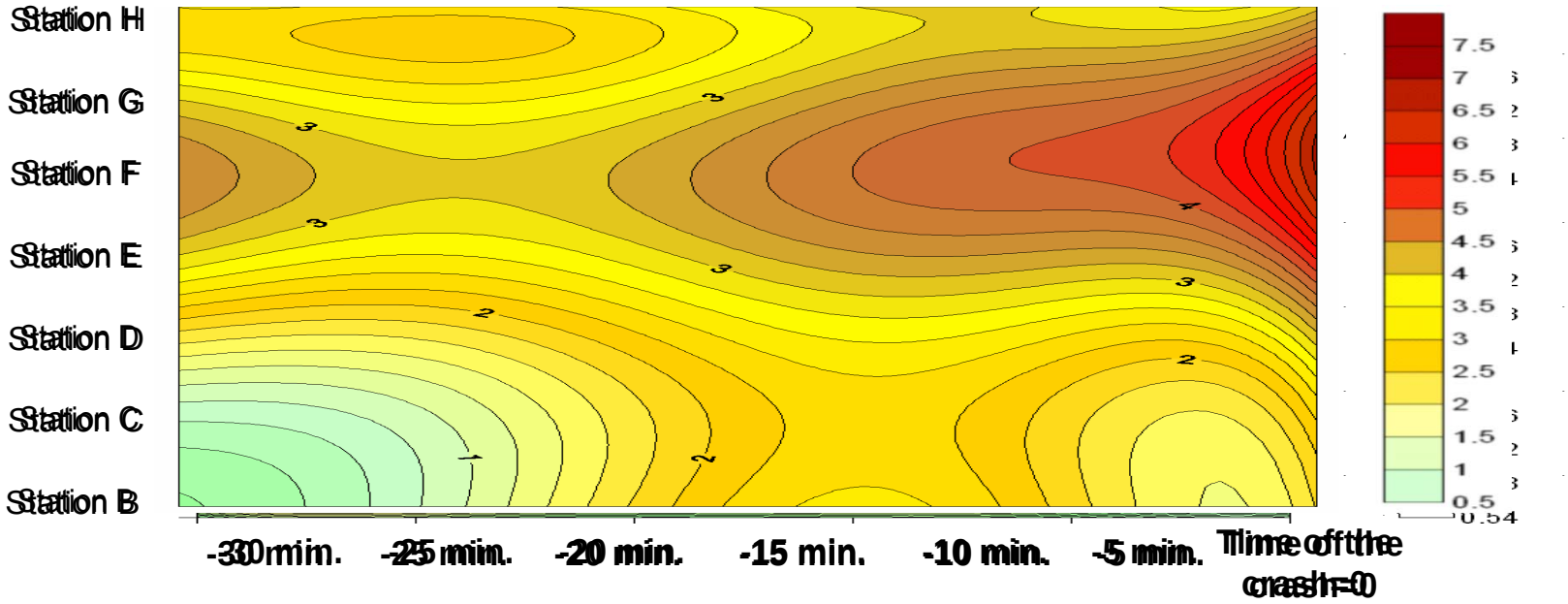
Big Data Applications in Microscopic Traffic Safety Analysis

- Proactive traffic safety management on I-4, Florida
  - Capturing the disturbance of traffic prior to crash occurrence in real-time
  - Developing a Hybrid Detailed Crash Prediction System Using ITS Data on I-4



Big Data Applications in Microscopic Traffic Safety Analysis

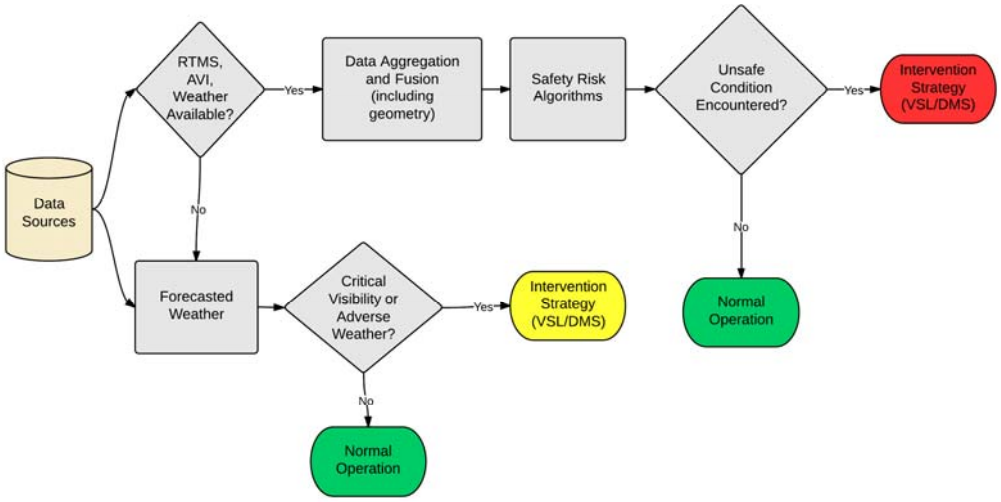
- Generating spatio-temporal crash risk plot for I-4



Hazard Ratio: Contour plots of hazard ratios corresponding to coefficient of variation in speed (42 model outputs)

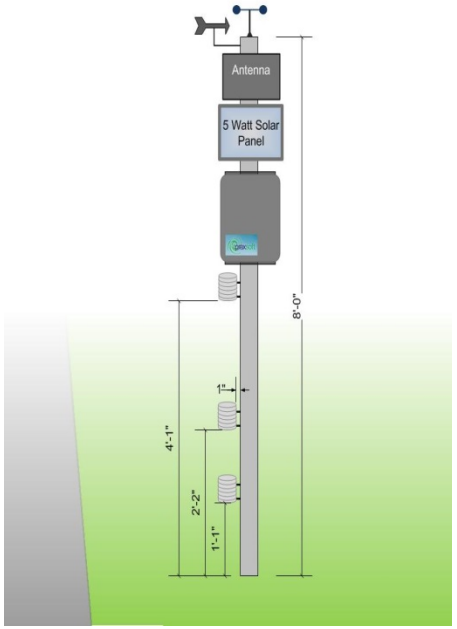
Big Data Applications in Microscopic Traffic Safety Analysis

- Developing an Active Traffic Safety Management System for I-70 in Colorado
  - Integration of Big Data from multiple sources
    - AVI, MVDS (aka RTMS), and weather detectors
- Big Data solutions to safety problems
  - Micro-simulation using VISSIM
    - Variable Speed Limit proved effective to mitigate speed variation



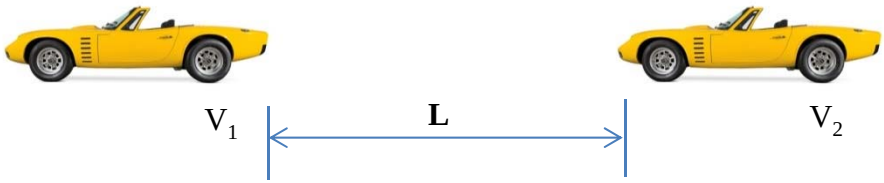
Big Data Applications in Microscopic Traffic Safety Analysis

- Big Data to capture previously unmeasured vehicle movement on I-4, Polk County



Big Data Applications in Microscopic Traffic Safety Analysis

- Generating surrogate measures of safety
  - Time to collision (TTC) under reduced visibility conditions
    - TTC=L/ (V_1-V_2)



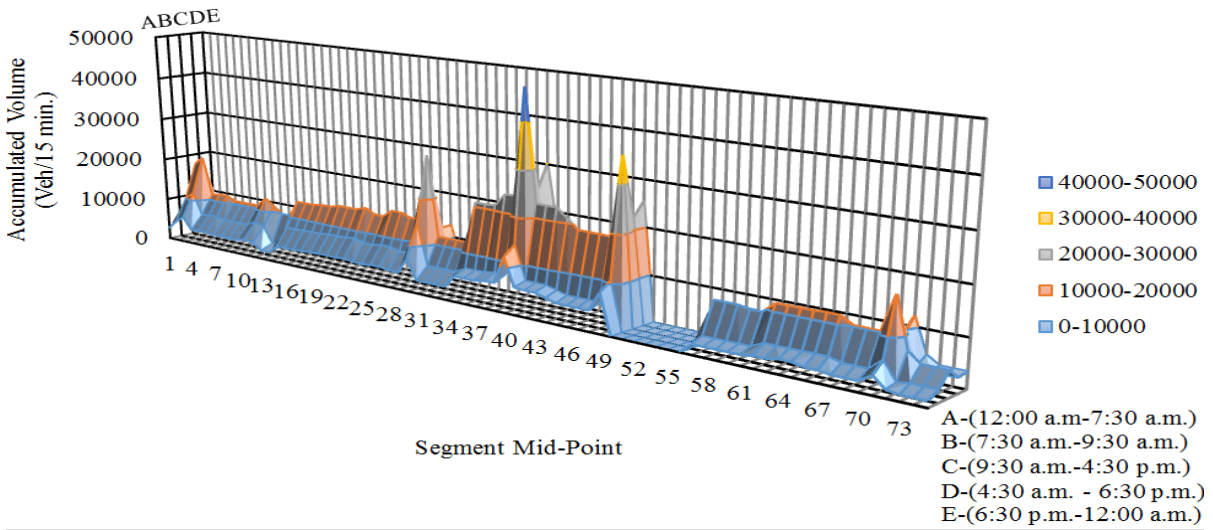
    - TTC_1 was calculated when V_1 maintained its own speed
    - TTC_2 was calculated when the leading vehicle suddenly stopped

	TTC_1			TTC_2		
Visibility change	Mean Difference	95% Confidence Interval		Mean Difference	95% Confidence Interval	
$\geq 2000m \rightarrow (100m - 2000m)$	7.36*	5.54	9.18	0.95*	0.83	1.07
$(100m - 2000m) \rightarrow \leq 100m$	18.24*	14.06	22.41	1.20*	0.93	1.47
$\geq 2000m \rightarrow \leq 100m$	25.60*	21.66	29.54	2.15*	1.90	2.41

    - Both TTC_1 and TTC_2 will decrease significantly as the visibility is reduced

Big Data Applications in Microscopic Traffic Safety Analysis

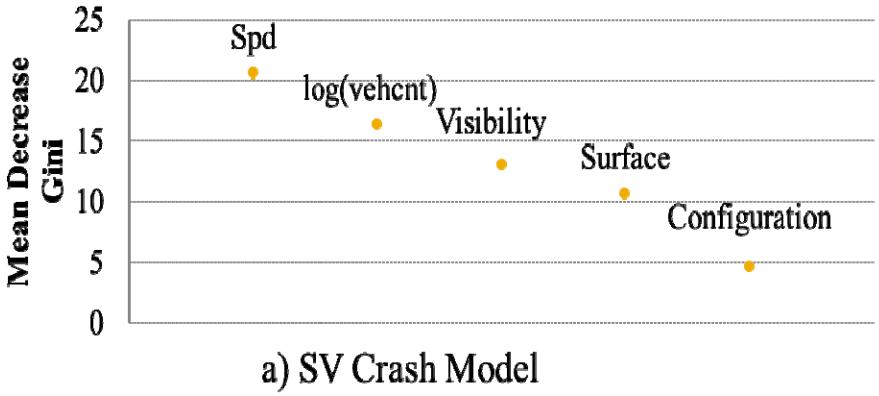
- Big Data for more dynamic safety screening on Central Florida Expressways
  - Fixed safety screening for the whole day period
  - Dynamic safety screening for different time period of day



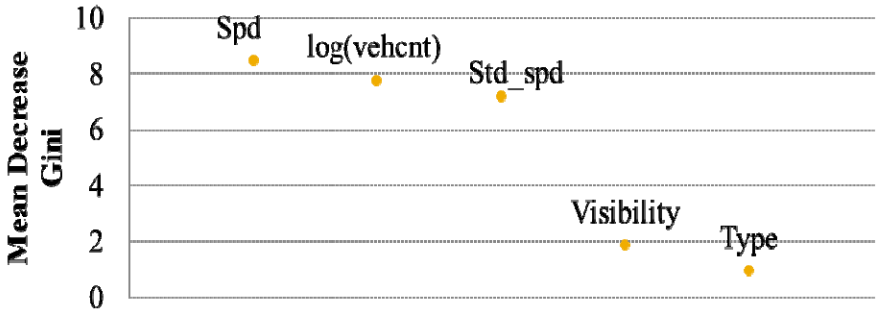
SR 408

Big Data Applications in Microscopic Traffic Safety Analysis

- Big Data analytics of real-time expressway ramp safety
  - Data mining: Random Forest
    - Importance ranking of crash contributing factors
    - Traffic >Weather >Geometry
    - Speed is the most important factor



a) SV Crash Model



b) MV Crash Model

Big Data Applications in Microscopic Traffic Safety Analysis

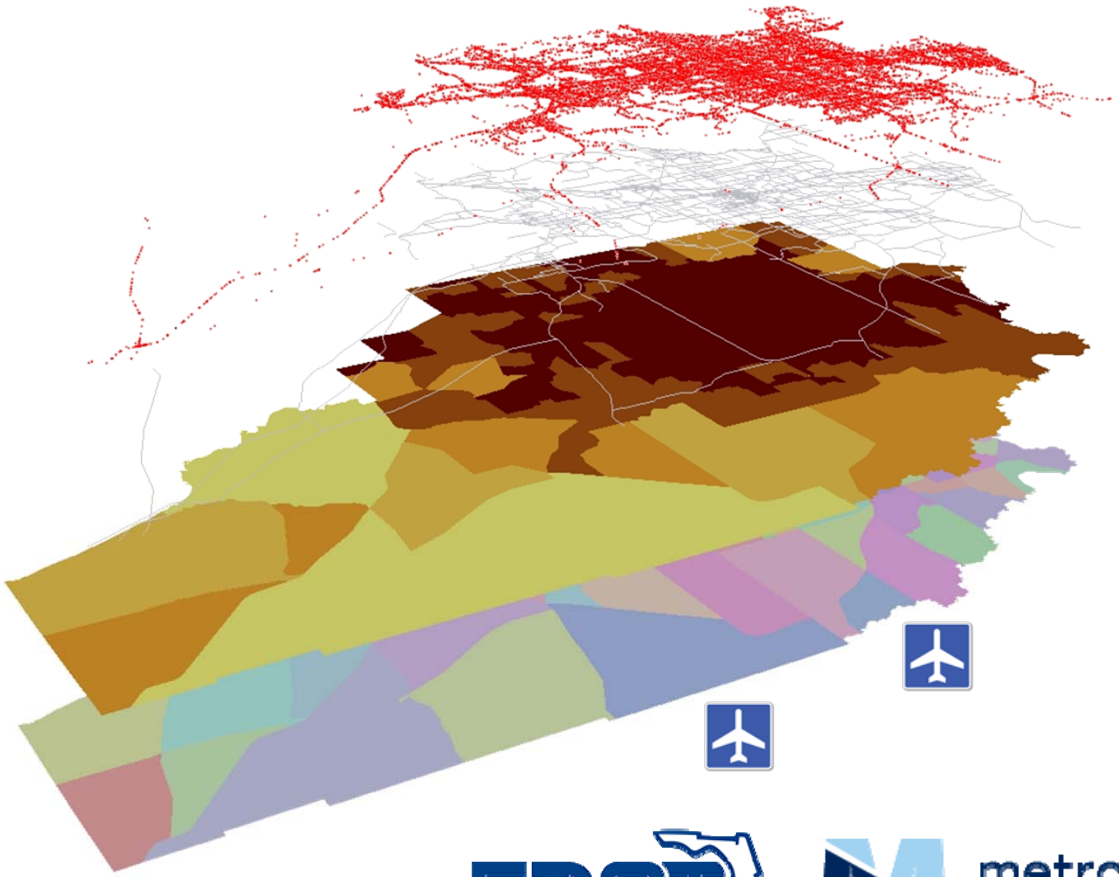
- Data-driven remedies to over-speeding on ramps
  - Speed Sensor
  - Colored DMS message

Overspeed Warning System on Loops

Normal condition

Overspeed condition

Multiple Data Sources



Layer 5: Crash data (State DOT)

Layer 4: Roadway data (State DOT)

Layer 3: Socio-demographic (US Census Bureau) and Land-use data (FDOR)

Layer 2: Traffic Analysis Zones (Regional MPO)

Layer 1: Weather information from adjacent weather stations (NOAA)

FDOT

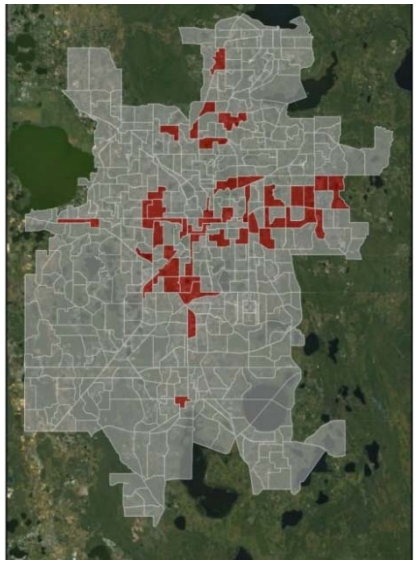
metroplan orlando
A REGIONAL TRANSPORTATION PARTNERSHIP

United States Census Bureau

NOAA

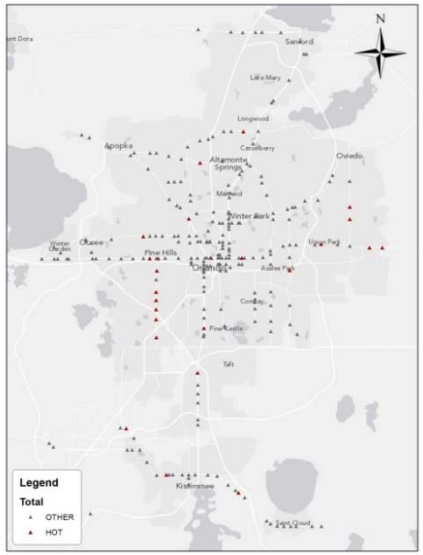
Big Data Applications in Macroscopic Traffic Safety Analysis

- Micro- and macro-level screening



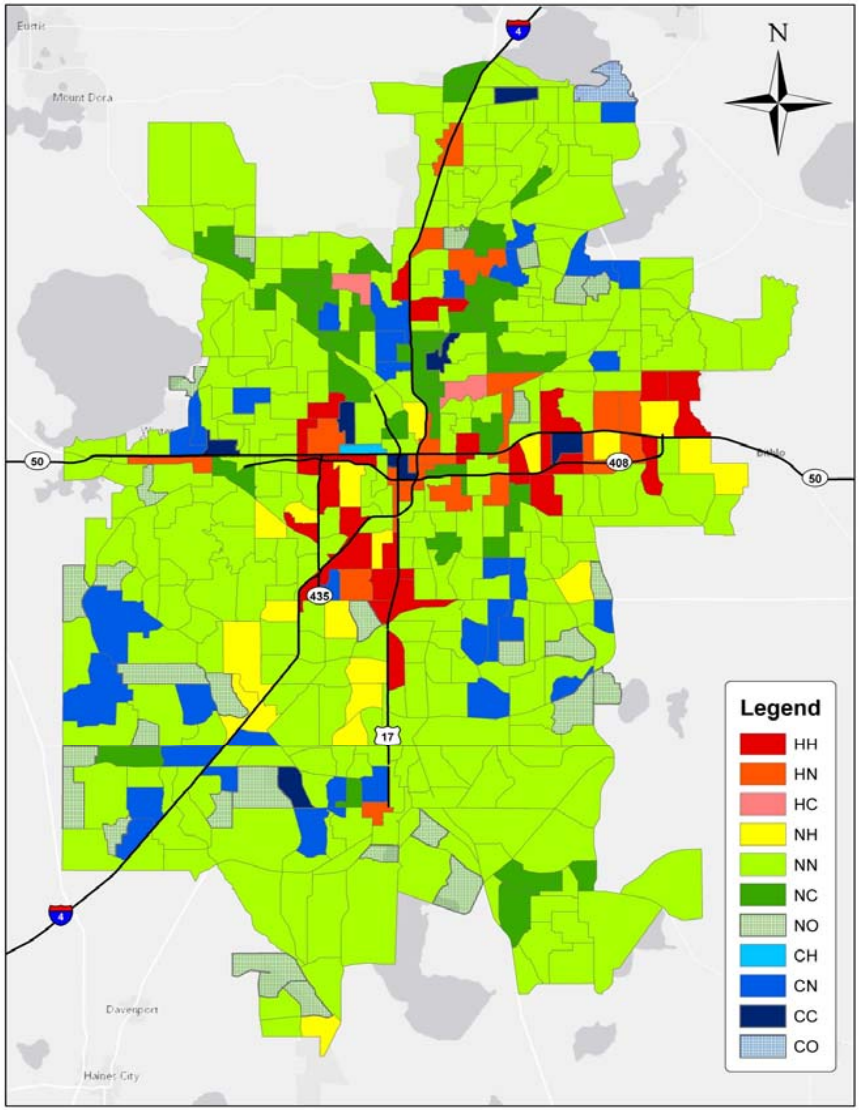
Macro-screening

+



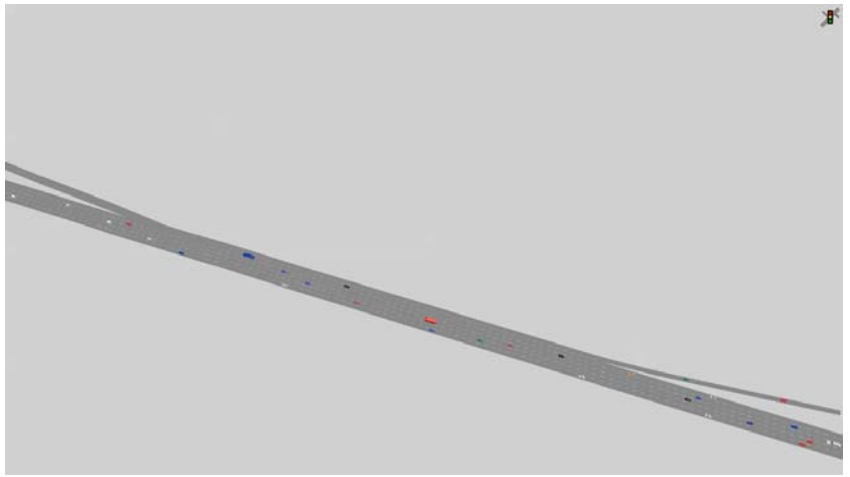
Micro-screening

Integration

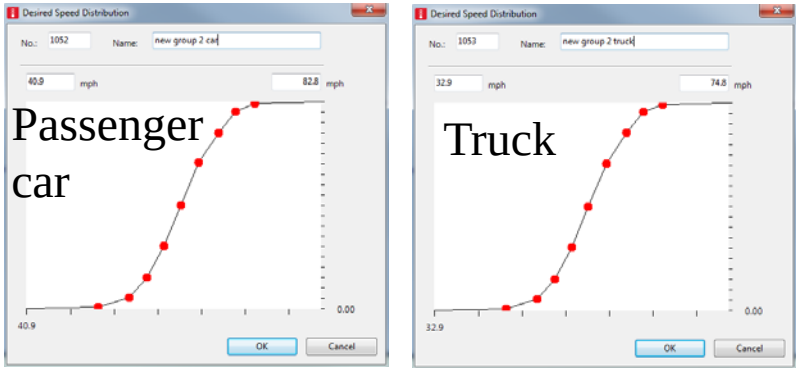


Other Potential Safety Evaluation Methods using Big Data

- Real-time traffic data for simulation
  - Input
    - Detailed and precise traffic input
    - Accurate field desired speed distribution
  - Calibration and validation result of 16 weaving segments
    - Simulated Volume: 96.0% consistent with the field volume
    - Simulated Speed: 92.2% consistent with the field speed
- Surrogate Safety Evaluation in Simulation
  - Simulated Conflict (SSAM)
  - Correlation with field crash is 0.513
- Conclusion
  - Big data can provide simulation networks which are highly consistent with field networks



SR408 WB Weaving Section @12.6-13.0

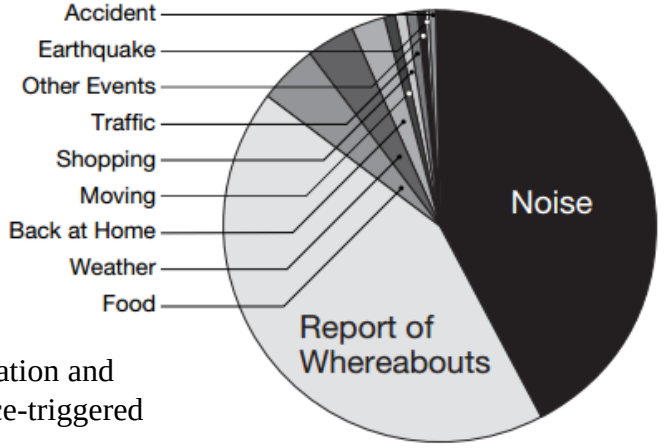


RTMS Speed distribution by vehicles types

Future Potentials of Big Data in Traffic Safety Management

- Access to new data
  - Social media on mobile devices
    - Large user base
      - Twitter: more than 300 million active users and 80% users on mobile
    - Rich information
    - Real-time and geo-tagged information

- Top Socially networked cities in the US
http://mashable.com/2011/03/19/top-100-cities/



*Data Source from Detection, Classification and visualization of place-triggered geotagged tweets

Common types of Tweets

Future Potentials of Big Data in Traffic Safety Management

- Event/incident detection
  - Providing more detailed information regarding the events
- Sentiment analysis of travelers' attitudes towards traffic safety and operation

- Creative approaches to summarize, describe and analyze the information contained in the data
- New applications of the data

Translation of tweet: Chaos on Capelseplein in the direction of Rotterdam, because of a delivery van in the crash barrier

*Data Source from
Twitter as a spatio-temporal information source for traffic incident management

Thank You

Dr. Mohamed Abdel-Aty, PE
Pegasus Professor and Chair
Department of Civil, Environmental & Construction Engineering

Deputy Director
Center for Advanced Transportation Systems Simulation
University of Central Florida, USA

Editor-in-Chief
Accident Analysis & Prevention

August 2015