UMTRI

A Multivariate Analysis of Crash and Naturalistic Event Data in Relation to Highway Factors Using the GIS Framework

4th SHRP2 Safety Research Symposium

Washington, DC
July 23 , 2009

T. Gordon, P. Green, L. Kostyniuk, M. Barnes,
A. Blankespoor, S. Bogard, D. Blower

UNIVERSITY OF MICHIGAN
TRANSPORTATION RESEARCH INSTITUTE

Objective

Capture common elements in how highway factors are associated with

- **Crashes (as recorded in crash data)**
- **Driving behaviors (as recorded in naturalistic driving data)**
- **Focus**
  - **On spatially referenced databases and GIS tools**
  - **On physical mechanisms, not just statistical association**

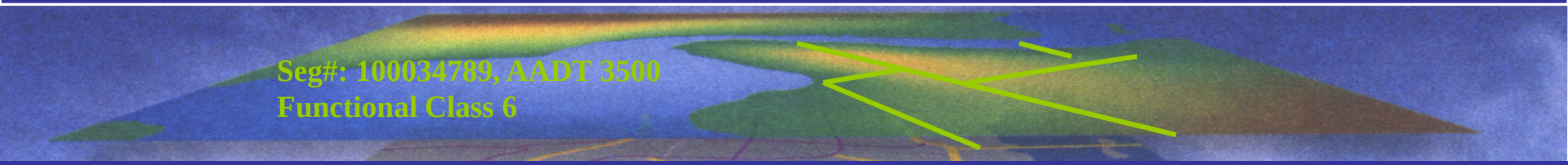
Approach

- Develop spatially-referenced database for road data, naturalistic driving data, and crash data
- Identify candidate surrogates related to disturbance of control
- Test validity against crash risk
  - Seemingly unrelated regression models – events
  - Models based on Extreme Value Theory – continuous
- Validate and replicate

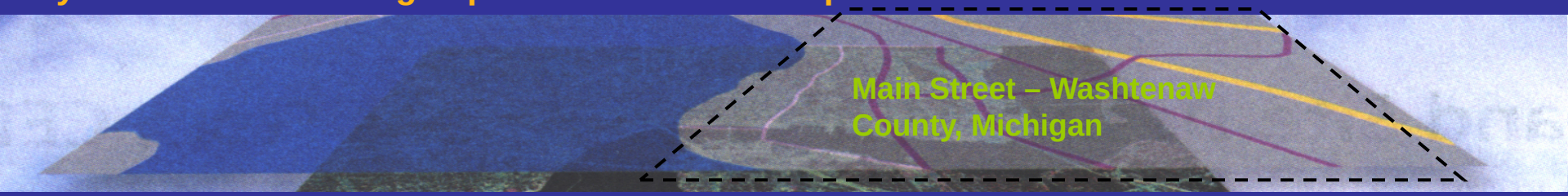
Spatially Referenced Data

Layer 4- SE Michigan Road Departure Crashes (2001-2005) (Lat/Long)

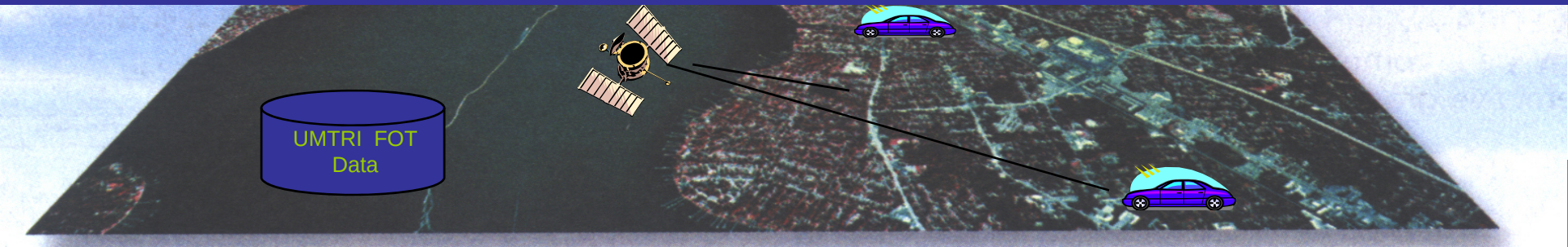
Layer 3 - Michigan 2005 HPMS + Direction - Table Data



Layer 2 - State of Michigan public roads base map



Layer 1 - UMTRI Naturalistic driving data from FOT fleets with Lat Long positions



Naturalistic Driving Data

- Road Departure Crash Warning (RDCW) Field Operational Test (FOT) - 2005
- 12 instrumented vehicles
- 78 subjects – 3 age groups, 2 genders
- 4-weeks of driving each
- 220,000 vehicle miles in SE MI 8-county area
- SQL server data base - raw vehicle data and derived tables (~250 Gb)
- Over 200 variables, including …
  - Differential GPS time and position
  - Lane tracking - includes boundary type (solid, dashed)
  - Distance to lane edge
  - Vehicle velocity
  - Yaw, pitch, roll
  - Front and side radar,
  - Face video
  - Front scene video
  - Sensor data on lights, windshield wipers

Crash Data

- **Road departure crashes from MI Vehicle Crash Data file 2001–2005**

- **71,308 crashes in SE MI 8-county area**
  - **70% on HPMS network**
  - **Direction of travel identified**
  - **Direction of road departure**
    - **15,187 left departures**
    - **23,836 right departures**
    - **32,287 unknown**

- **Driver - age, sex**
- **Environmental conditions (day/night, wet/dry pavement)**

Road Data

Our Universe - HPMS road segments with direction - traversed by FOT vehicles in SE MI 8-county area – (9,526 directional segments)

- **From HPMS**
  - Segment length
  - AADT
  - Functional class
  - Area type – Urban, Rural
  - Number of thru lanes
  - Shoulder type*
  - Shoulder width right*
  - Shoulder width left*
  - Speed limit*

*Sample data – involves imputation

- **From FOT data**
  - Curve presence
  - Average radius of curvature for segment
  - Average change of heading for segment
  - Available lateral maneuvering distance

Candidate Surrogates
from FOT data

- Time to lane crossing
- Time to road edge crossing (TTEC)
- Imminent lane departure warning (LDW Alert)
- Approach speed too fast for curve (curve speed warning)
- Lateral deviation in lane (LDEV)
- Peak projected lateral deviation
- Driver looking away for 2 sec+
- Coherency between tracking error and vehicle/driver response (Yaw Rate Error)
- Steering rate less than small threshold for at least 4 seconds
- Others from search for benchmarks of disturbed control
  - Cell phone use
  - Yaw deviation associated with discontinuity in lane markings

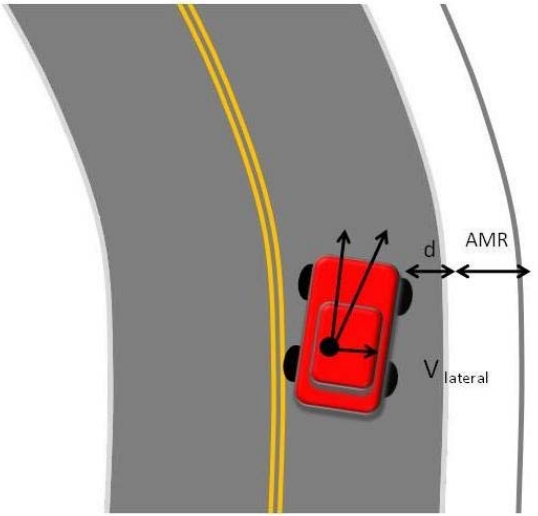
Candidate Surrogate Time to Edge Crossing (TTEC)

- TTEC = distance to outside edge of shoulder divided by lateral velocity

- Conditions:
  - Lane tracking confidence above 70%
  - Vehicle velocity in direction of solid boundary > 0 m/sec
  - No turn signal

- Variables for analysis:
  - Categorical - Counts of TTEC events < 5th percentile TTEC on road segment
  - Continuous – Minimum TTEC for each subject under set of conditions

$$TTEC=(d+AMR)/V_{lateral}$$



UMTRI

Candidate Surrogate
LDW Alert

- **Predicted vehicle path will cross solid lane boundary (edge or center line)**
  - **Conditions:**
    - **Vehicle speed >25 mph**
    - **Restriction on high steering rate in past 5 seconds**
    - **No turn signal in the past 5 sec**
    - **No brake in the past 5 sec**
    - **Not on local street**
    - **Actual tracking on boundary to be crossed**
  - **Variable for analysis**
    - **Categorical – Counts of LDW alerts for directional road segment**

Need analysis method to account for covariance of crashes and surrogates

Crash Data

Response	Crash Events
	(Exposure)
Explanatory Factors	Driver
	Environment
	Vehicle
	Highway

Naturalistic Driving Data

Response	Surrogate Events
	(Exposure)
Explanatory Factors	Driver
	Environment
	Vehicle
	Highway

Seemingly Unrelated Regressions (SUR) Zellner (1962)

$$Y_1 = X_1\beta_1 + \varepsilon_1 \qquad \text{crashes}$$
$$Y_2 = X_2\beta_2 + \varepsilon_2 \qquad \text{surrogates}$$

As a system of equations, SUR may be written as

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

crashes

surrogates

Or as a new linear model

$$Y = X\beta + \varepsilon$$

$$\varepsilon \sim N(0, \Sigma) \longleftarrow \text{Random component}$$

$$Var(\varepsilon) = \Sigma = \begin{bmatrix} \sigma_{11}I & \sigma_{12}I \\ \sigma_{21}I & \sigma_{22}I \end{bmatrix} \longleftarrow \text{Correlation structure}$$

Properties of SUR

1) The system satisfies the properties of Weighted Least Squares (WLS)

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$$

2) A correlation structure is induced between Y_1 and Y_2

3) Tests of hypotheses may be conducted

$$H_0 : \beta_1 = \beta_2$$

Bayesian SUR for Log-Linear Models
Estimate model parameters and distributions

Likelihood:

(After suitable transformations)

$$Y_1 \sim \text{Normal}(X_1\beta_1, \Sigma_1)$$ **crashes**

$$Y_2 \sim \text{Normal}(X_2\beta_2, \Sigma_2)$$ **surrogates**

Or as a system: $$Y \sim \text{Normal}(X\beta, \Sigma)$$

Vague Prior:

$$\beta \sim N_p(0, 10^6 I)$$

$$\Sigma$$ **Fixed covariance matrix estimated from ordinary least squares (OLS) residuals**

Relative Risks

- Can obtain Relative Risk (RR) from model results
- Can use RR as exposure-based risk measure
- Test if there is a difference in RR between crashes and surrogates for levels of an explanatory variable, holding other variables constant
- If log RR difference is zero, conclude that the candidate surrogate has basic validity

Surrogate Candidate -TTEC

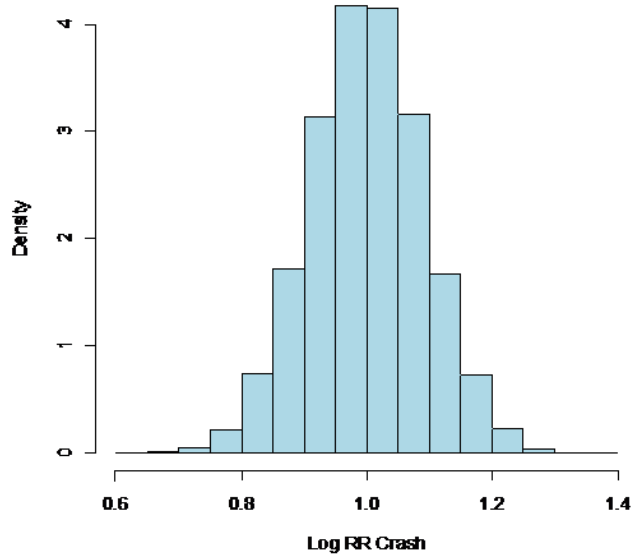
- **Response Variables**
  - Number of road departure crashes to the right
  - Number of right TTEC events
- **Exposure**
  - For crashes - AADT x 5 x segment length
  - For surrogate - number of FOT traversals in directional segment x segment length
- **Explanatory Variables**
  - Curve — Yes, No
  - Freeway — Yes, No
  - Area — Rural, Urban
  - Right shoulder width — 0-3ft, 3^{+}-8 ft, >8 ft

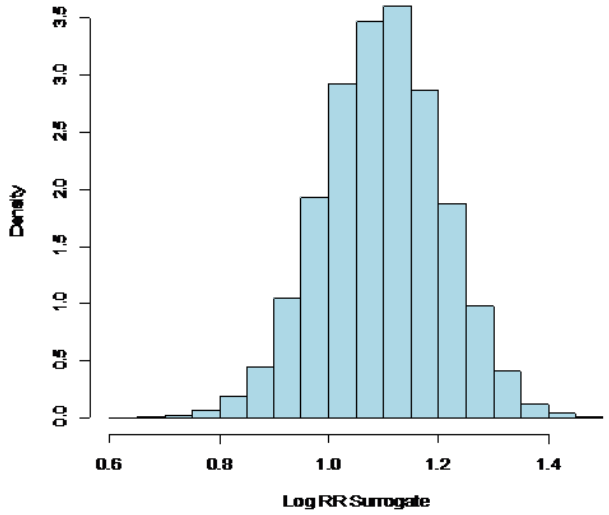
Relative Risks from Bayesian SUR Log Linear Model

Histograms of 30,000 Random Variables Simulated from the Posterior Distributions of the Log RR Crash and Log RR Surrogate (TTEC)

Comparing: Curve to No Curve

Not On Freeway, In Rural Area
Shoulder Width 3^{+}-8 ft





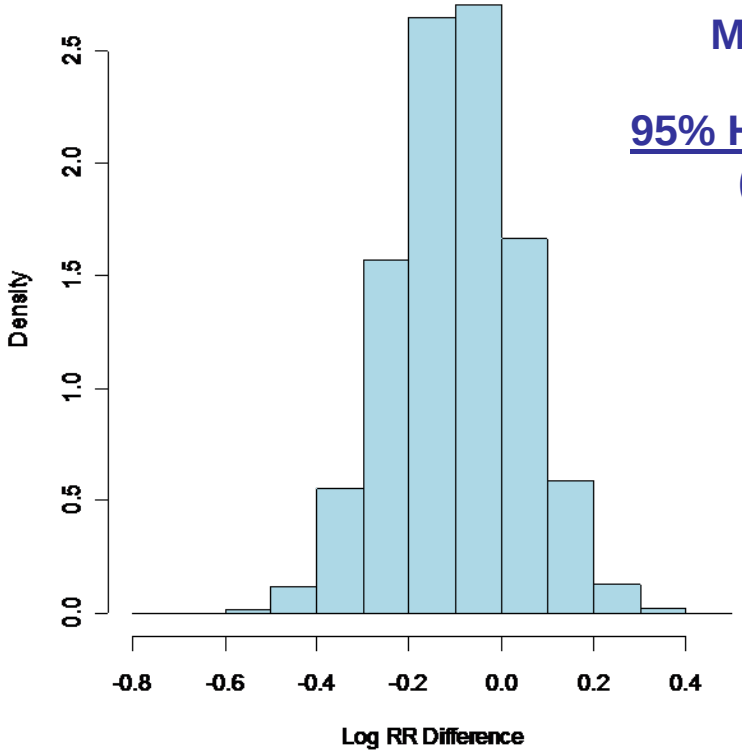
UMTRI

Crashes and TTEC
Estimated Difference in Log RR

(Histogram of 30,000 Random Variables Simulated from the Posterior Distribution of the Log RR Difference)

Comparing: Curve to No Curve

Not On Freeway
In Rural Area
Shoulder Width 3-8 ft



Mean: - 0.10

95% HPD Credible Set
(-0.36, 0.17)

UMTRI

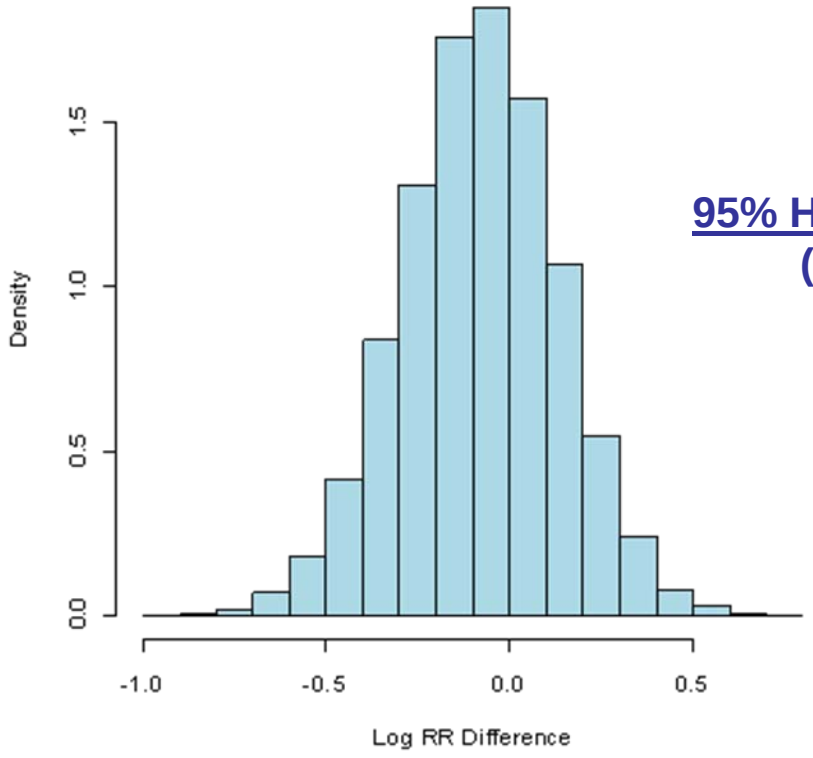
Crashes and LDW Alerts
Estimated Difference in Log RR

(Histogram of 30,000 Random Variables Simulated from the Posterior Distribution of the Log RR Difference)

Comparing:

Curve to
No Curve

Not On Freeway
In Rural Area
Shoulder Width 3-8 ft



Mean: -0.08

95% HPD Credible Set
(-0.51, 0.33)

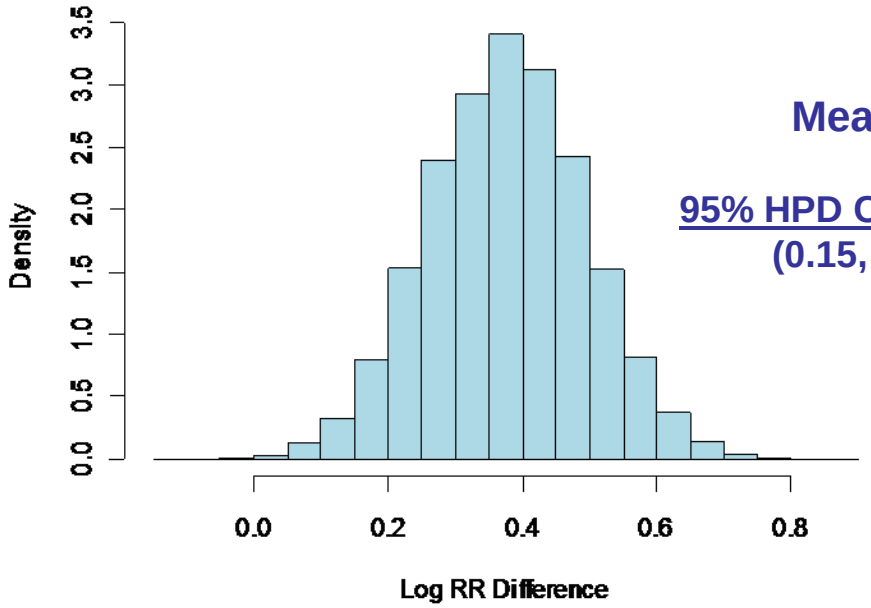
UMTRI

Crashes and LDEV
Estimated Difference in Log RR

Histogram of 30,000 Random Variables Simulated from the Posterior Distribution of the Log RR Difference)

Comparing: Curve to No Curve

Not On Freeway
In Rural Area
Shoulder Width 3-8 ft

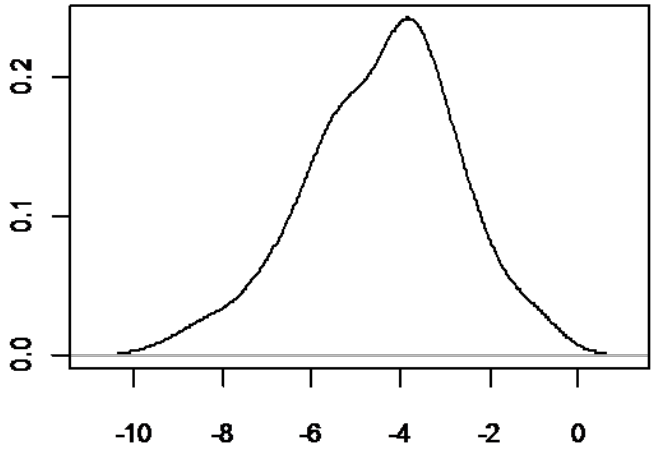


Mean: 0.38

95% HPD Credible Set
(0.15, 0.61)

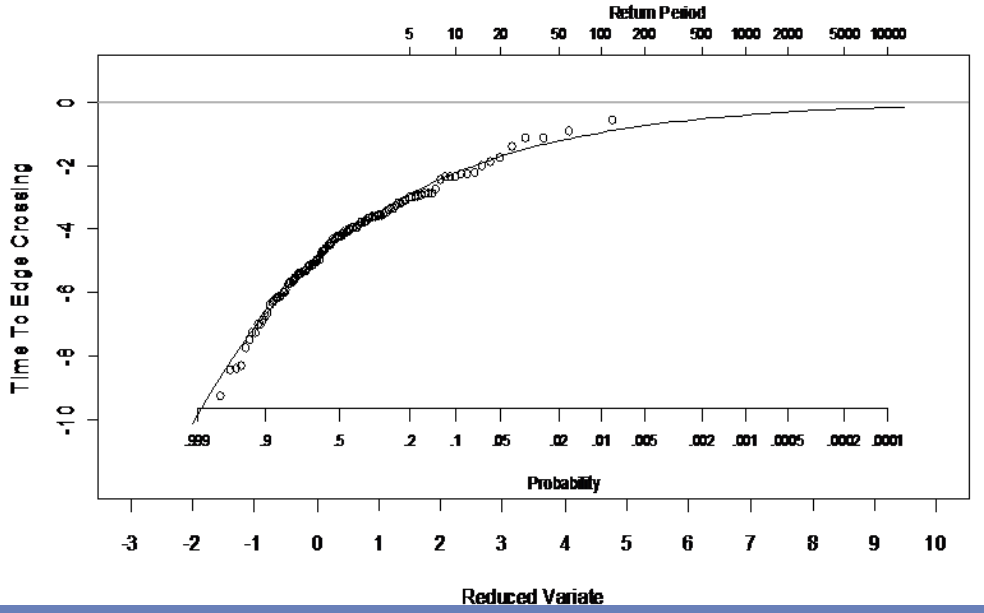
UMTRI

Second analysis method using extreme value theory to model rare events that lie outside the range of available observations



Kernel density of min TTEC
117 observations
2.3 miles – freeway near UMTRI

Weibull – crosses 0 at about 2 x10^6



UMTRI

Crashes

TTEC Extremes

AADT=65,755 veh

Expect a road departure about 12 times/year

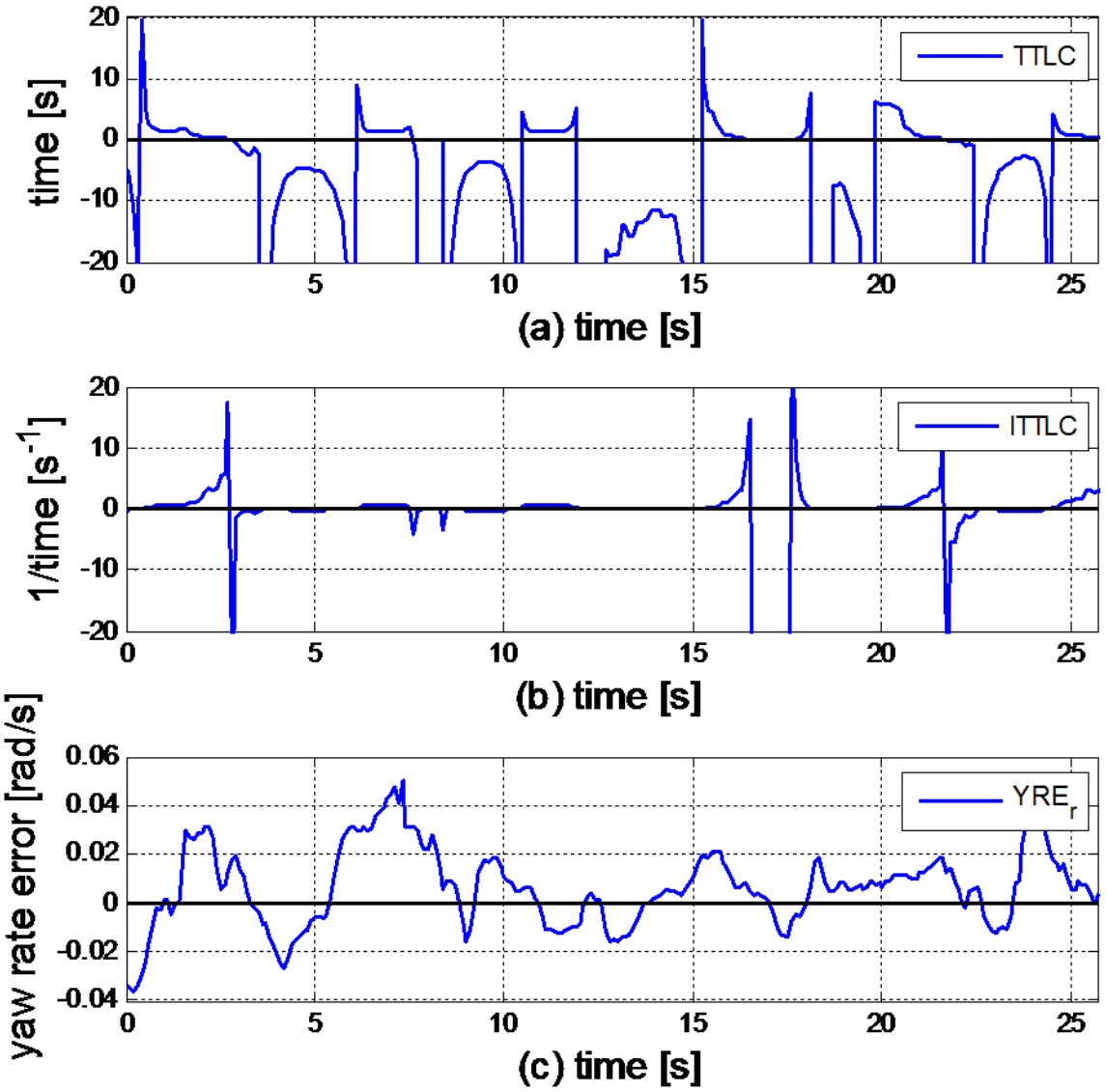
Crashes/year = 1.8

About 15% of road departures result in crash on this road

UMTRI

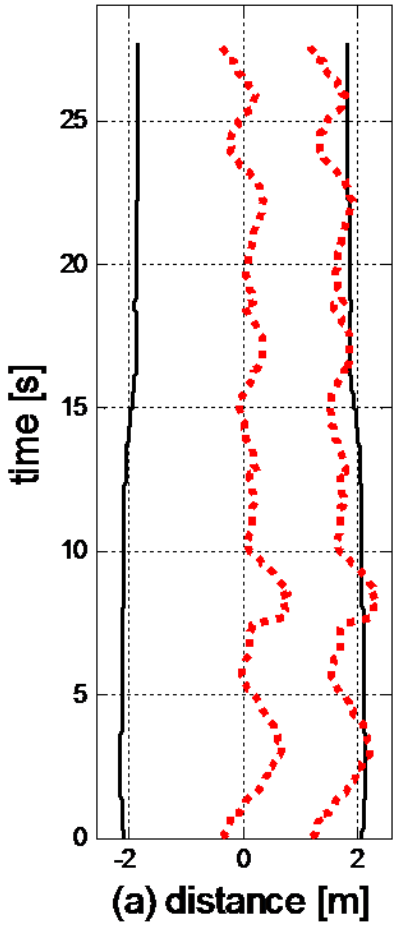
- Higher level surrogate – yaw rate error
  - Locate lane boundary points
  - Calculate critical yaw rate for each instance
  - Compute limits based on critical boundary points
  - Predictive and more sensitive than lane position
  - More continuous than lane TTLC, TTEC …

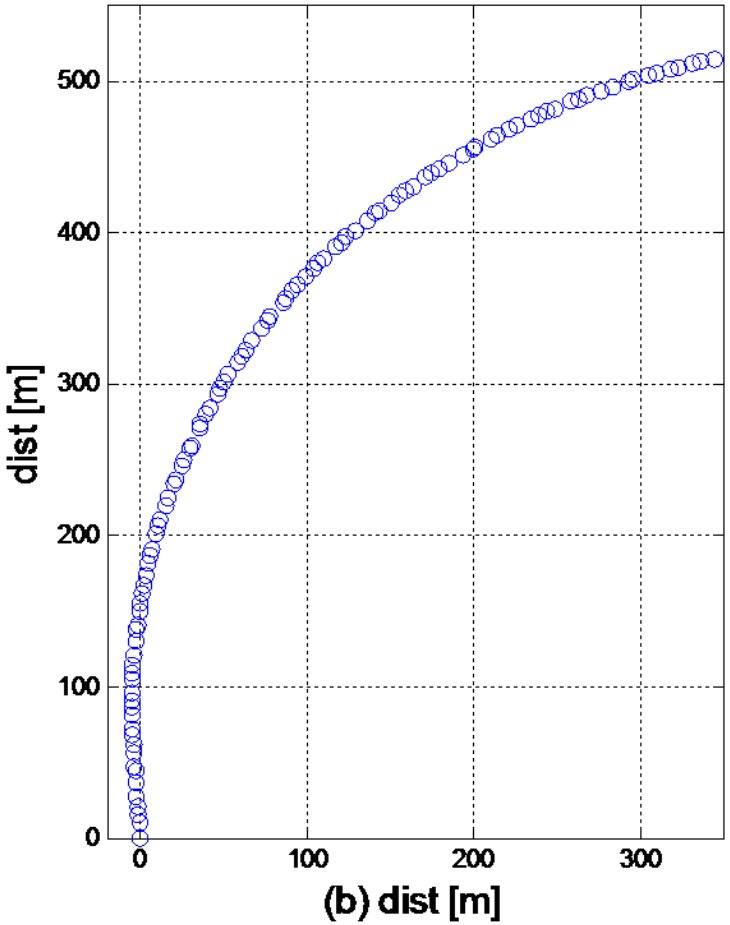
Event 1



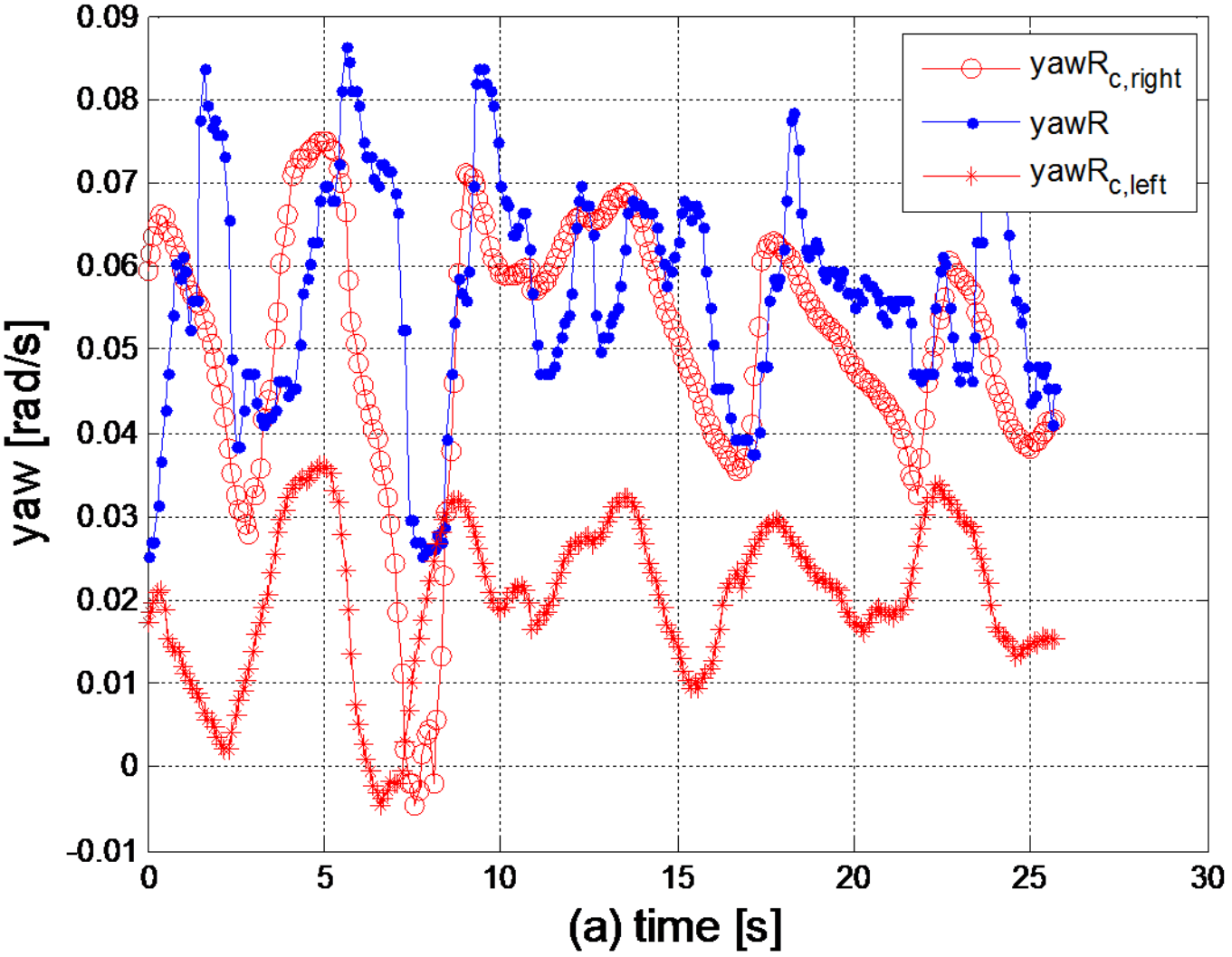
UMTRI

Event 1

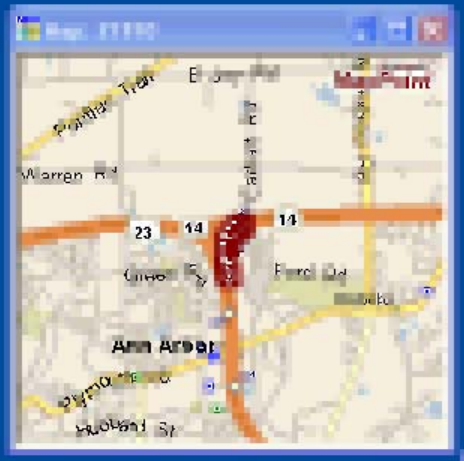

time [s]
0
5
10
15
20
25
-2
0
2
(a) distance [m]

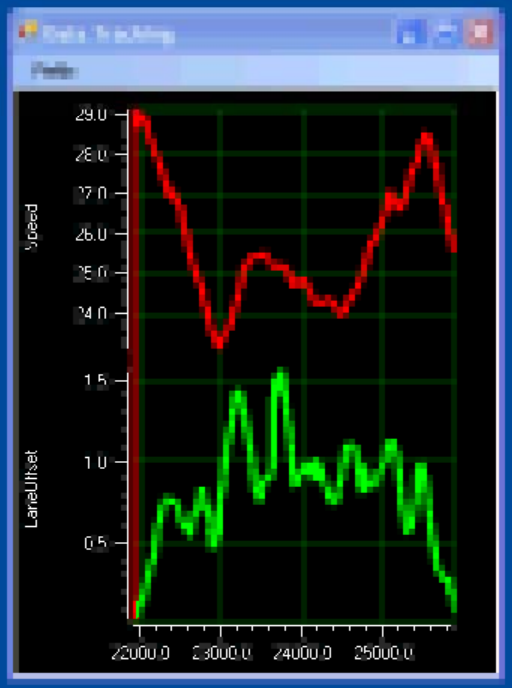

dist [m]
0
100
200
300
400
500
0
100
200
300
(b) dist [m]

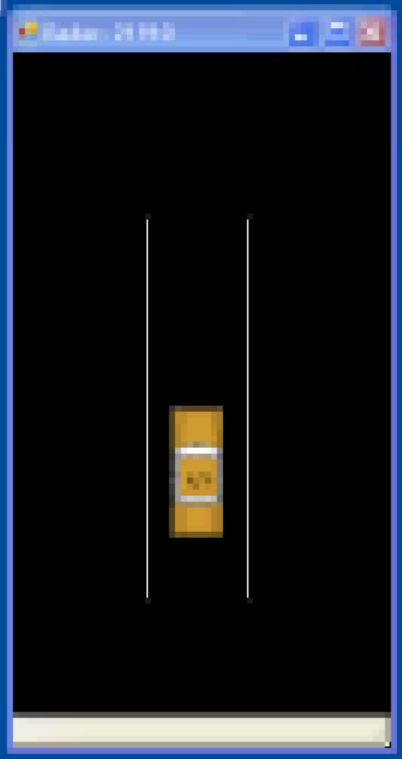
Event 1



UMTRI



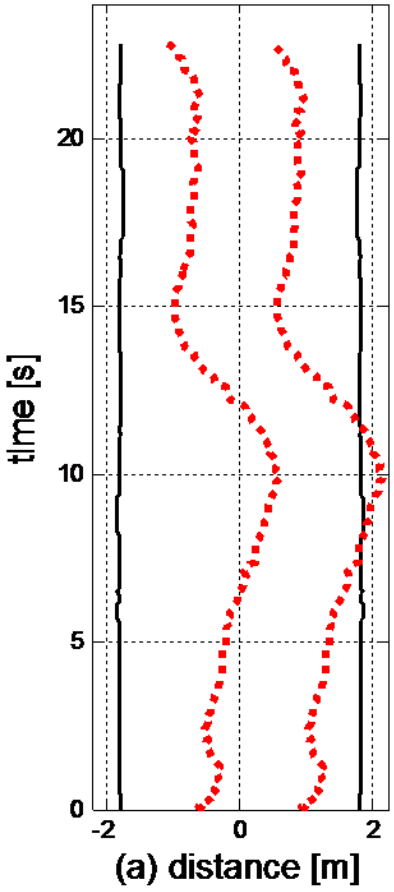


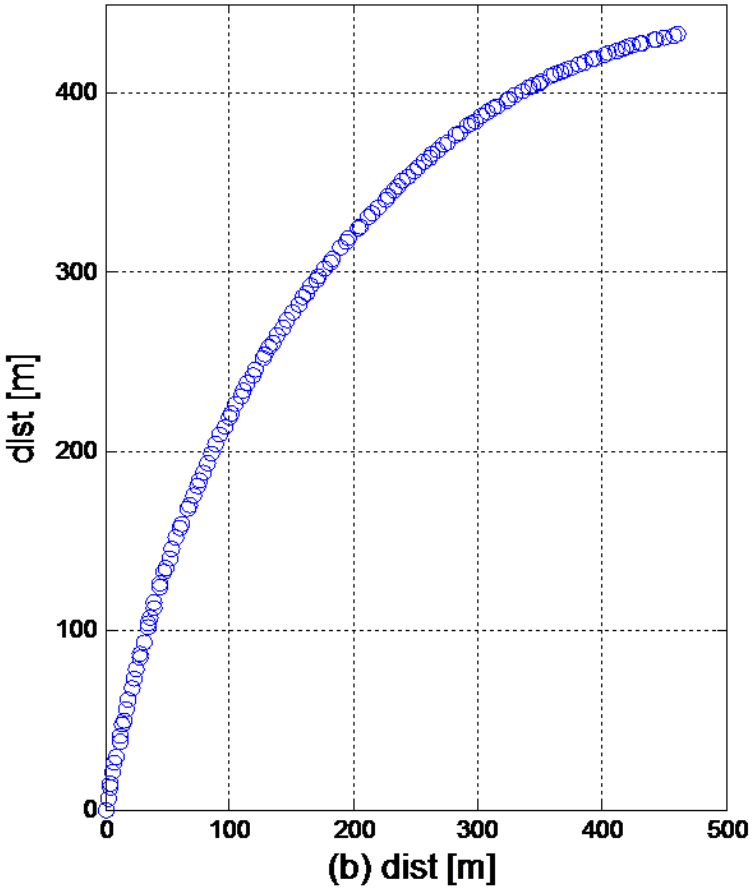


Next video at time 21960.

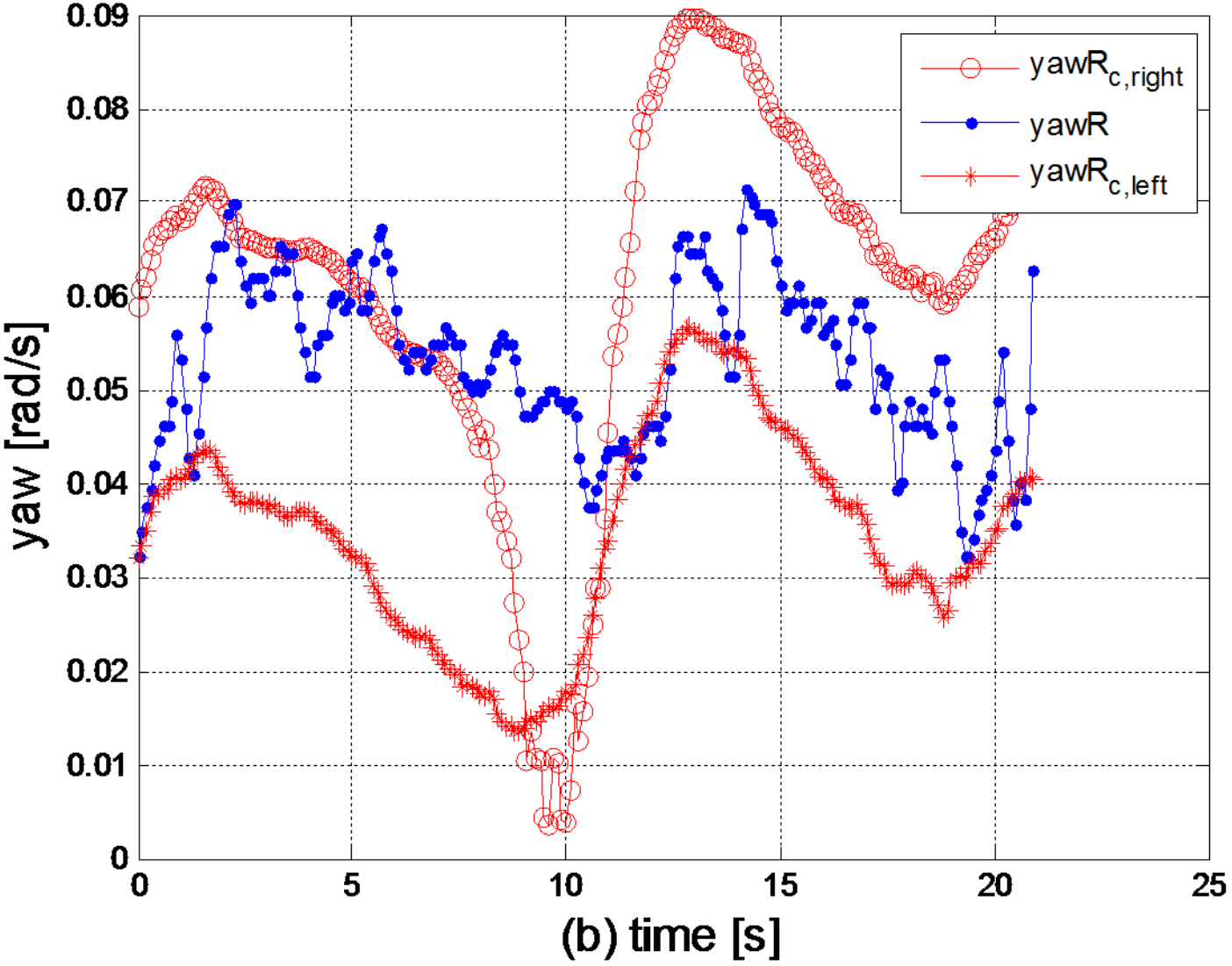
Event 1

Event 2

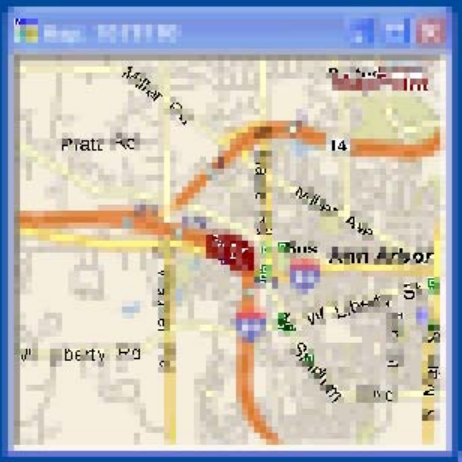


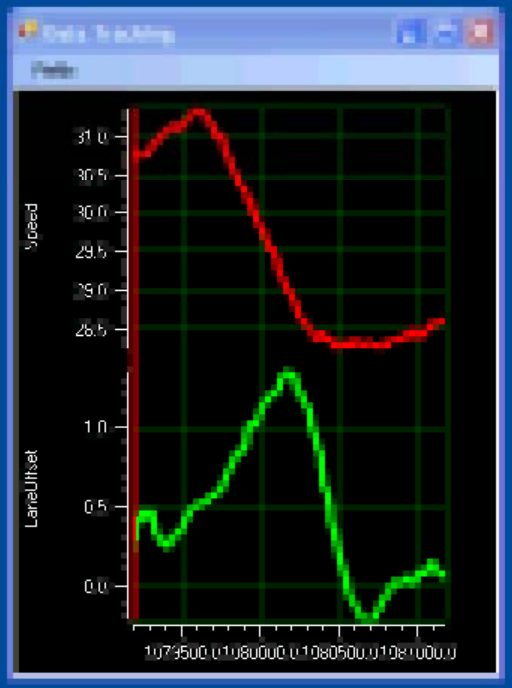


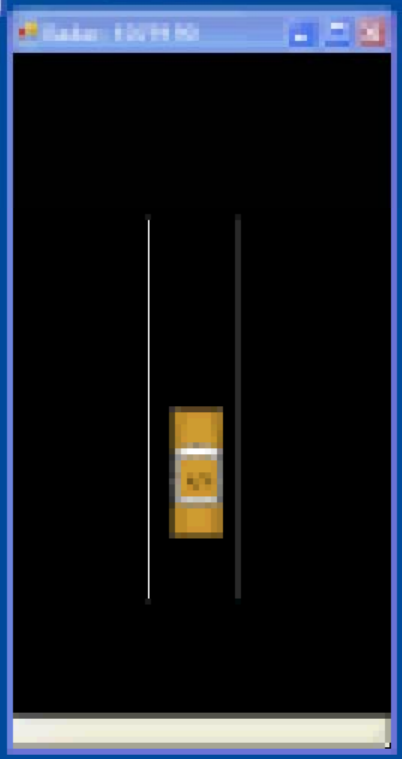
UMTRI


0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
yaw [rad/s]
0
5
10
15
20
25
(b) time [s]
yawR$_{c,right}$
yawR
yawR$_{c,left}$

UMTRI

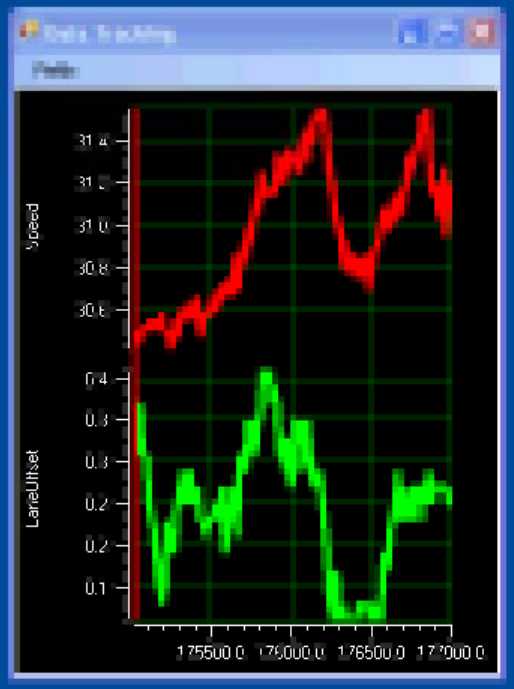


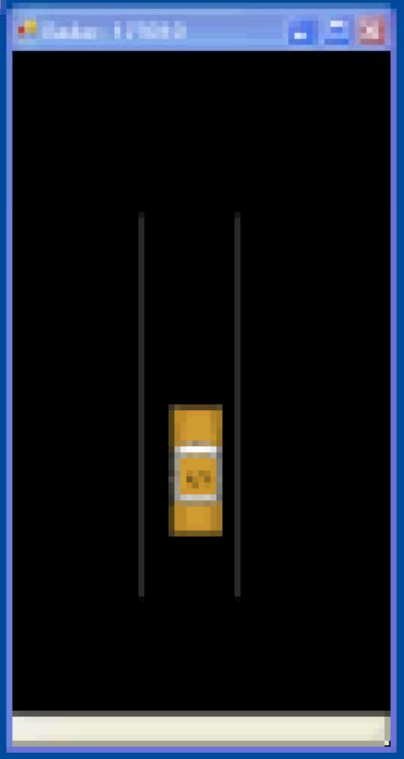




Next video at time 1079270.

Event 2

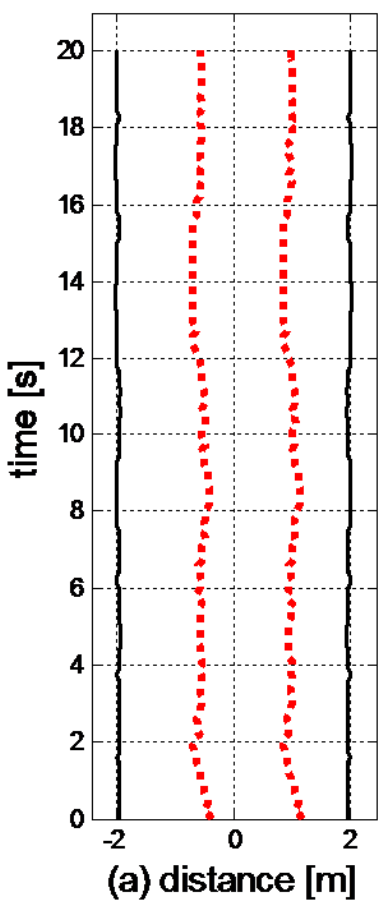


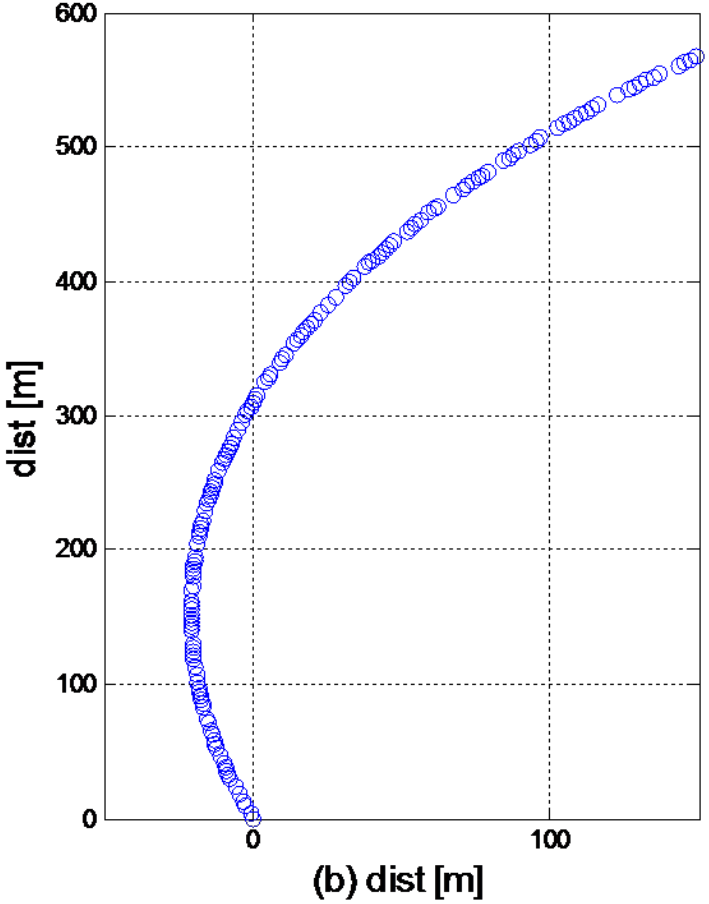


Next video at time 175080.

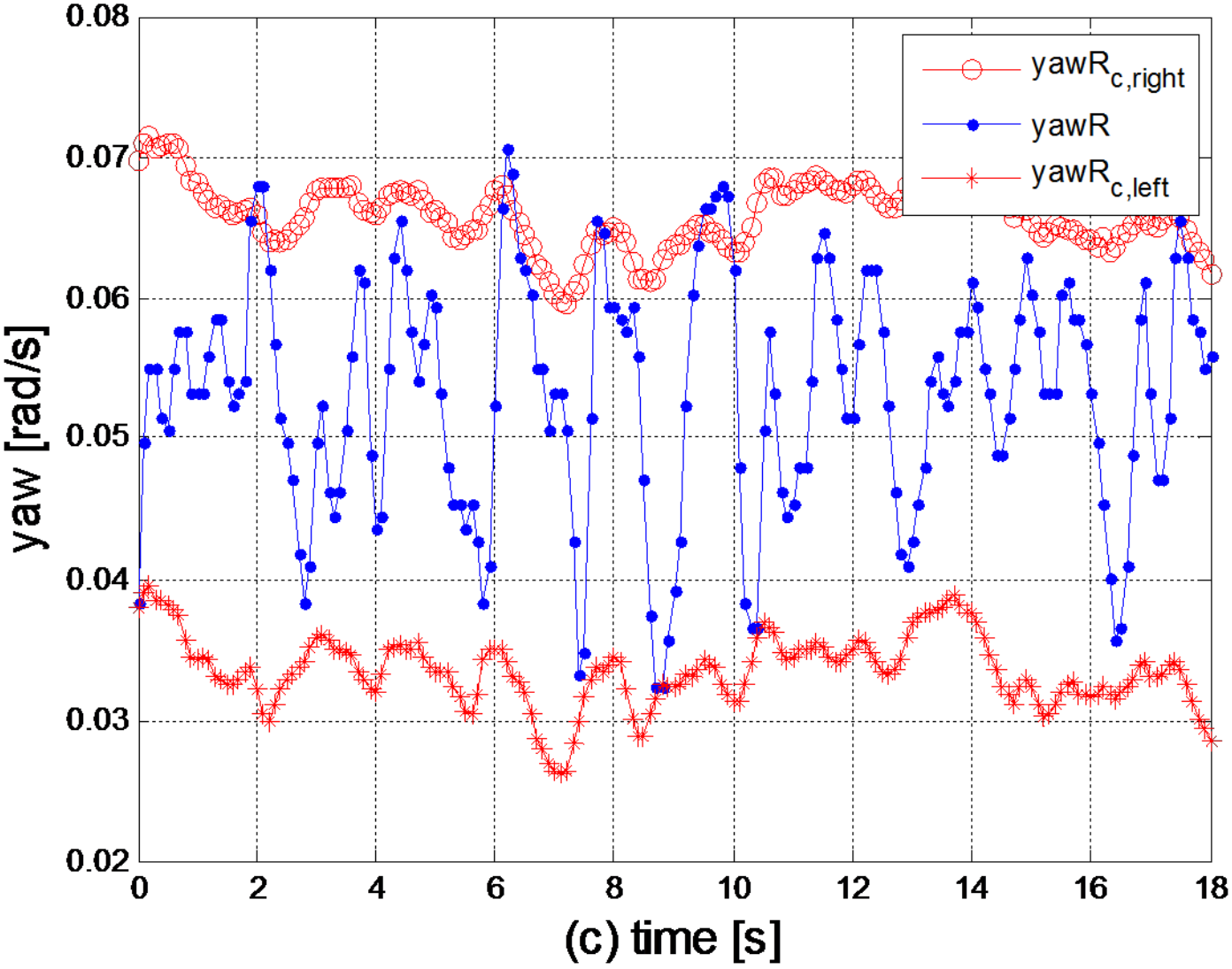
Event 3

Event 3





UMTRI


0.08
0.07
0.06
0.05
0.04
0.03
0.02
yaw [rad/s]
0
2
4
6
8
10
12
14
16
18
(c) time [s]
yawR$_{c,right}$
yawR
yawR$_{c,left}$

UMTRI

Thank You !