Discriminant Analyses

Jeremy Sudweeks
VTTI

Driving Transportation with Technology

VirginiaTech Transportation Institute

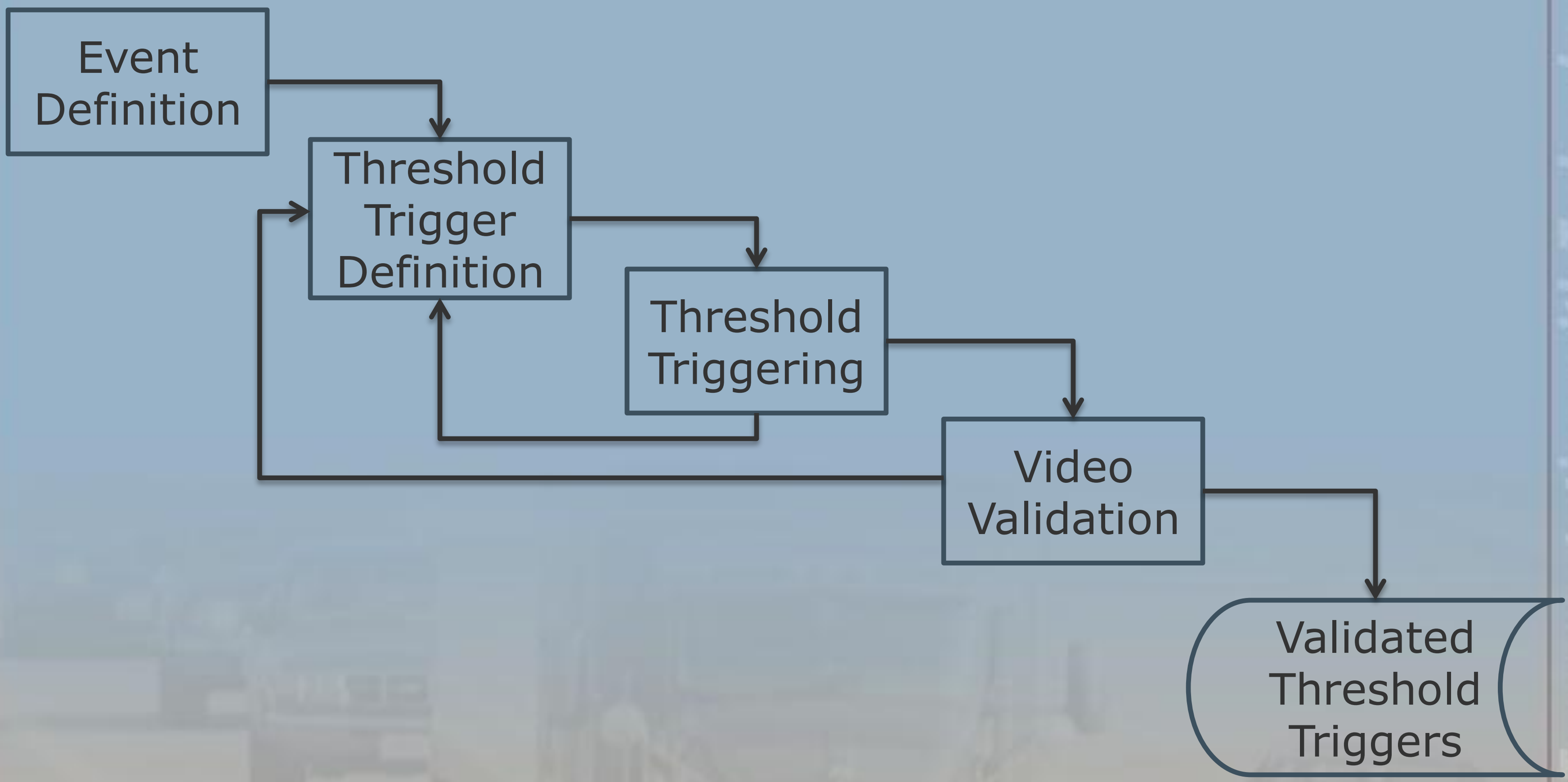
Objectives

- Discuss aspects of naturalistic data important for consideration during classification efforts
- Discuss established and recently developed classification methods amenable to naturalistic data

Outline

- Discuss threshold trigger distribution and yaw raw trigger criteria
- Graphical exploratory analysis of yaw threshold triggers
- Review linear discriminant analysis
- Brief introduction to high dimensional classification

Trigger Validation



100 Car Threshold Triggers

Invalid Threshold Triggers

Threshold Trigger	Trigger Frequency	Percent Invalid
Forward TTC	25,536	19.3
Lane abort	1,206	0.9
Lane solid	1,209	0.9
Lateral accel.	3,269	2.47
Lon. accel.	7,937	6.00
Rear TTC	866	0.66
Side blind spot	1,507	1.14
Side blinker	3,845	2.91
Side cutoff	853	0.64
Side yaw	1,396	1.06
Yaw rate	84,648	64.00

Valid Threshold Triggers

Threshold Trigger	Trigger Frequency	Percent Valid
Forward TTC	5,371	46.63
Lane abort	2	0.02
Lane solid	8	0.07
Lateral accel.	88	0.76
Lon. accel.	3,675	31.91
Rear TTC	440	3.81
Side blind spot	4	0.03
Side blinker	3	0.03
Side cutoff	261	2.27
Side yaw	15	0.13
Yaw rate	1,651	14.33

100 Car Threshold Triggers

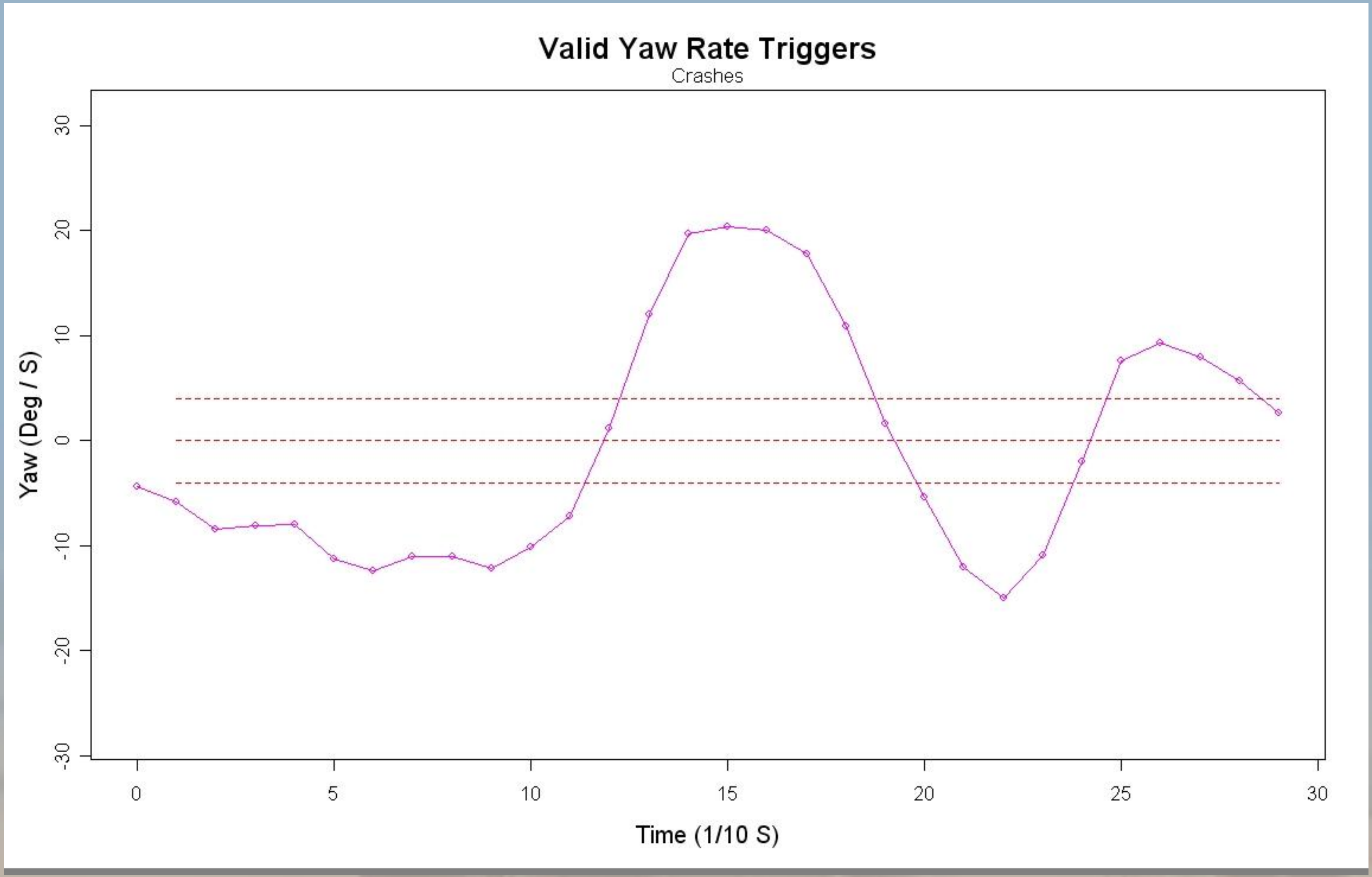
Threshold Trigger	Event Frequency	Percent Event
Forward TTC	2,827	33.2
Lon. Accel., Forward TTC	1,933	22.7
Yaw Rate	1,452	17.1
Lon. Accel.	1,348	15.8
Side Cutoff	255	3
Forward TTC, Rear TTC	129	1.5
Lon. Accel., Forward TTC, Rear TTC	125	1.5
Rear TTC	104	1.2
Other	336	3.95

Yaw Rate Threshold Trigger Definition

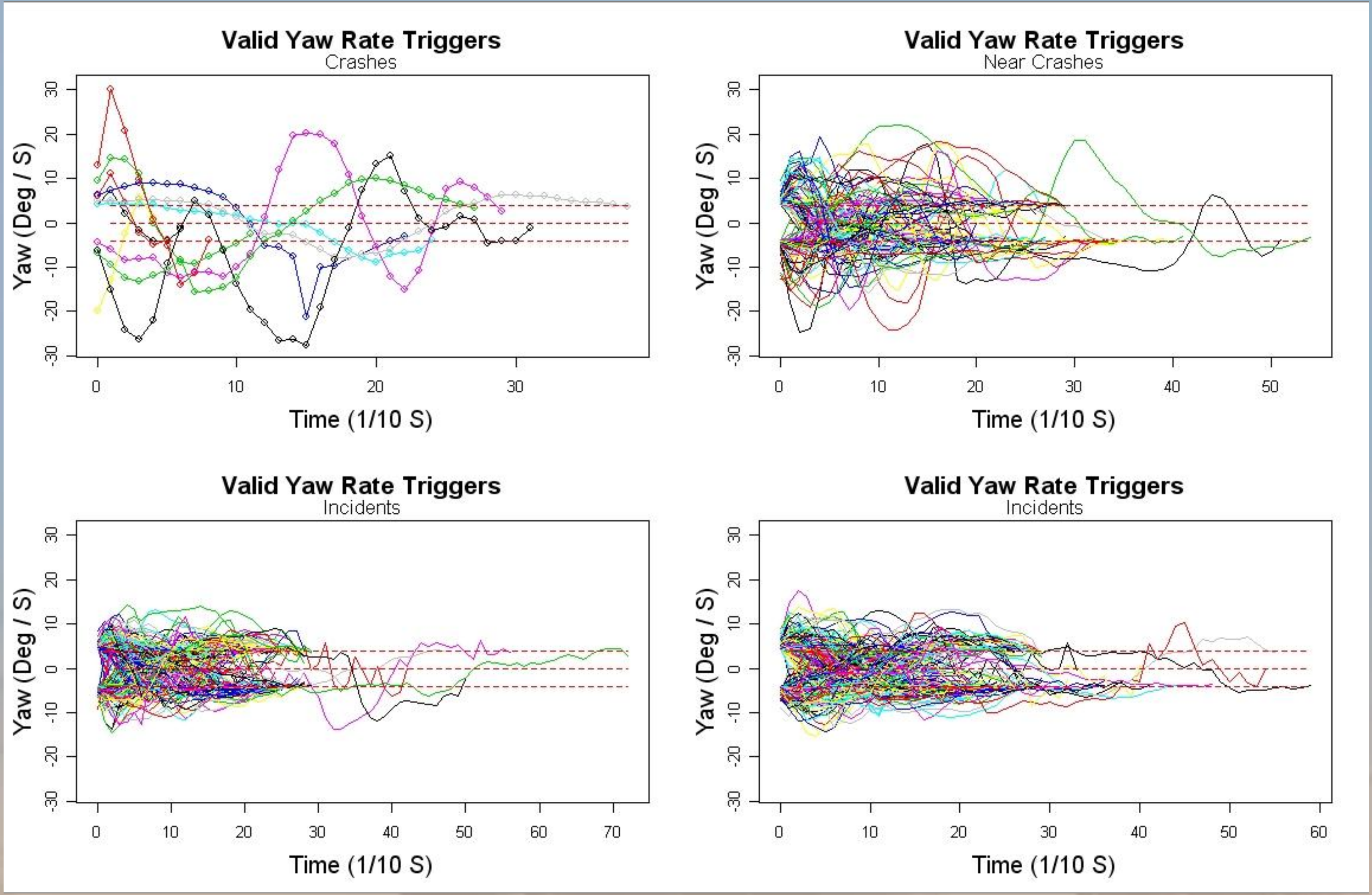
- The purpose of this threshold trigger was to identify situations in which a driver performed a sudden steering maneuver

- The final trigger criteria was as follows:
  - Yaw rate oscillation in excess of 4 degrees/second within a 3 second window (vehicle returned to direction of travel prior to steering manuever)
  - A minimum speed of 6.7 m/s (15 mph) at the onset of the trigger

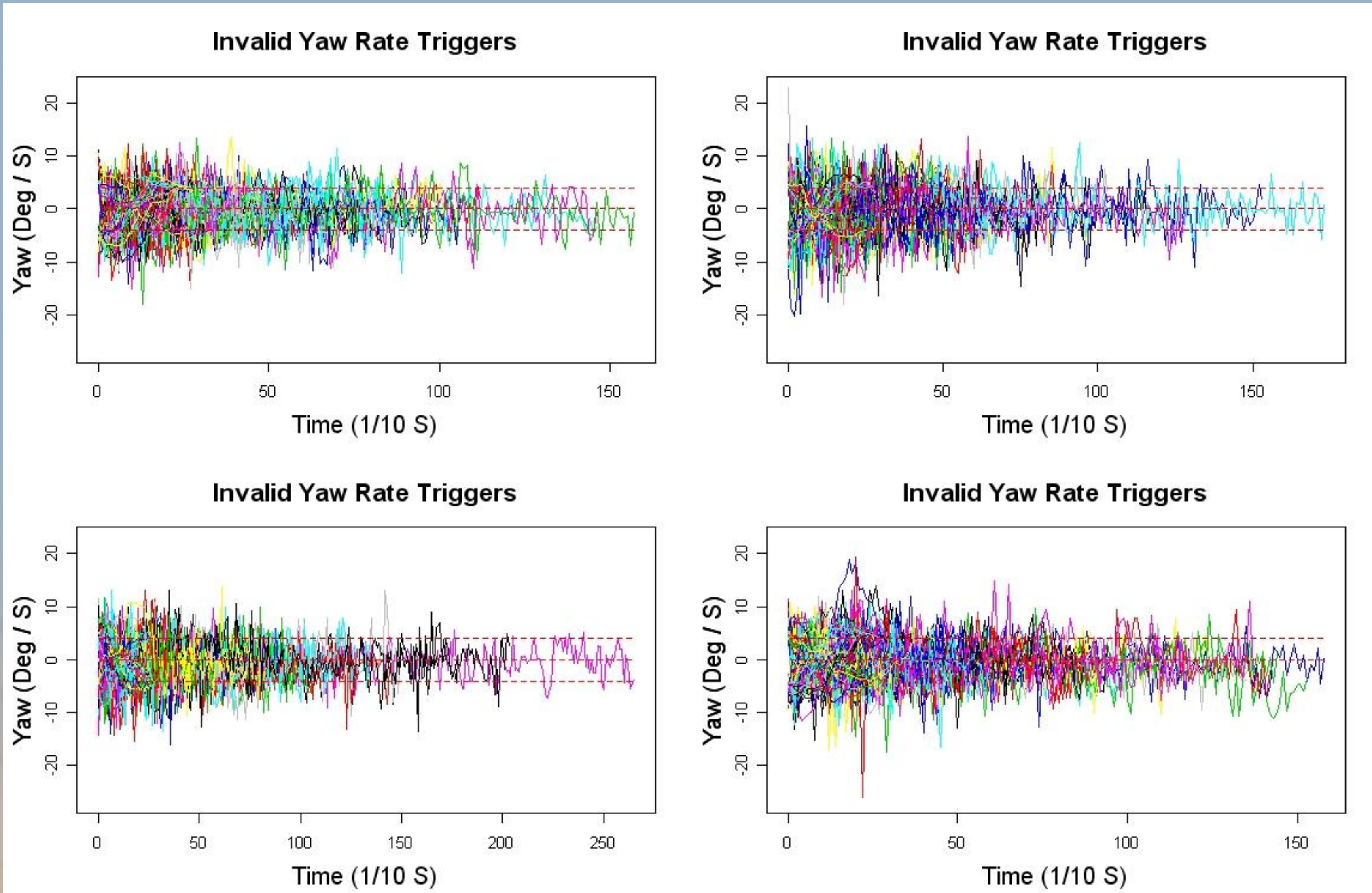
Yaw Rate Profile Plot



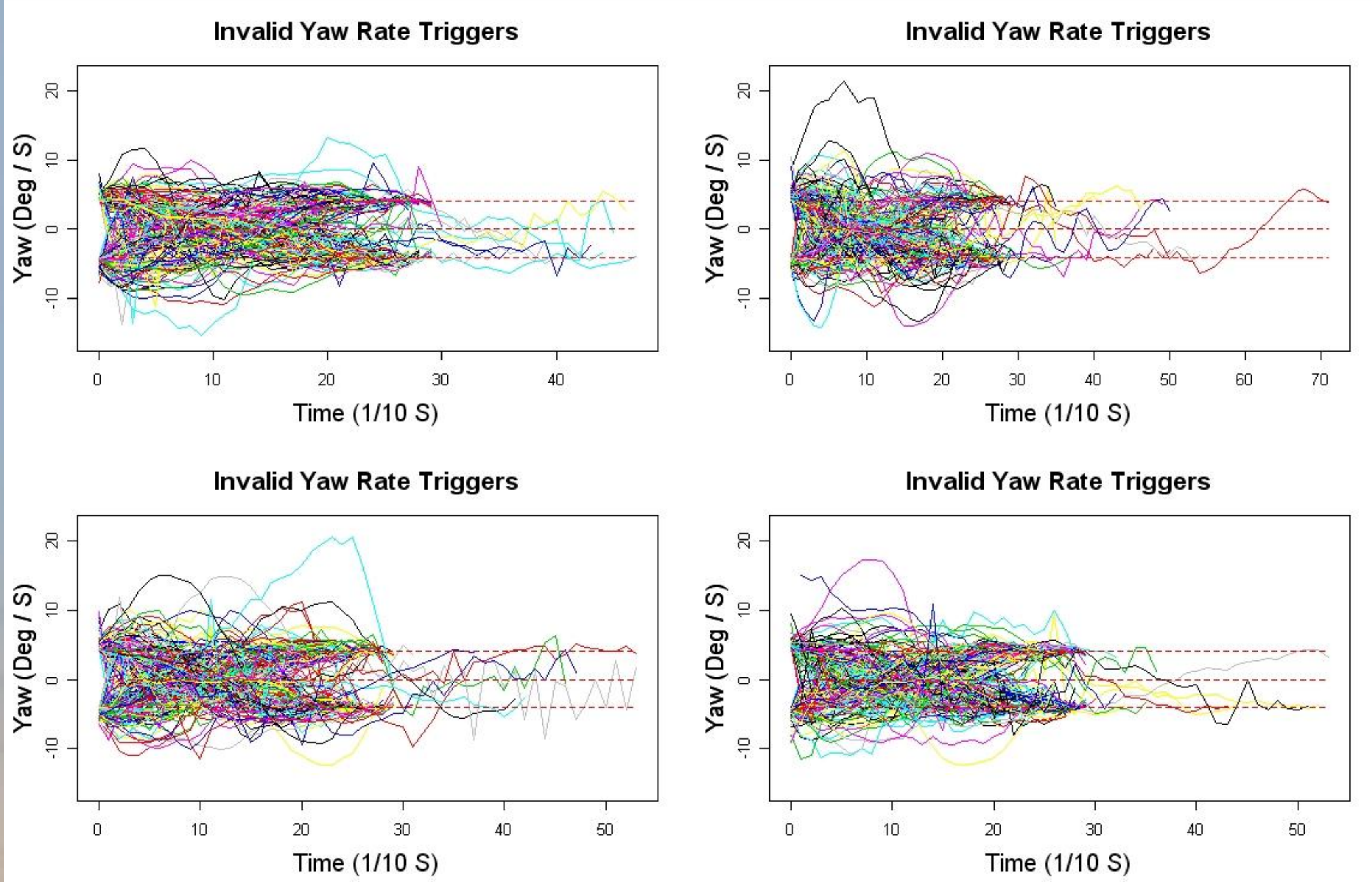
Yaw Rate Profile Plot



Yaw Rate Profile Plot



Yaw Rate Profile Plot



Summary of Exploratory Analysis

- Event definitions are difficult to fully specify
- Instrumentation noise and error need to be taken into account
- Class imbalance – valid yaw rate threshold triggers are rare
- Repeated observations of individuals
- Threshold trigger correspondence to onset of event is unknown
- Profile characterization is difficult

Classification Task

- Based on information contained in samples drawn from two populations (G_1 & G_2) create a decision rule to classify new observation vector **y**

- In the case of yaw rate threshold trigger profile characterization can be difficult

Classification Methods

- Statistical
  - Linear discriminant analysis (LDA)
  - Quadratic discrminant analysis (QDA)
  - Regularized discriminant analysis (RDA)
  - Kernel based methods
  - Nearest neighbor
  - Multivariate adaptive regression splines (MARS)
  - Classification and regression trees (CART)
  - Random forests
  - Functional discriminant analysis
- Neural networks
- Machine learning
  - Support vector machines
- Formal taxonomies for classification methods exist (Holmstrom, et al 1997)

Linear Discriminant Analysis

- Assumptions
  - $\Sigma_1 = \Sigma_2$

- A classification rule attributed to Fisher:
  - Assign **y** to G_1 if

$$a'y = (\bar{y}_1 - \bar{y}_2)' S_{pl}^{-1} y > \frac{1}{2} (\bar{y}_1 - \bar{y}_2)' S_{pl}^{-1} (\bar{y}_1 + \bar{y}_2)$$

  - And assign **y** to G_2 if

$$a'y = (\bar{y}_1 - \bar{y}_2)' S_{pl}^{-1} y < \frac{1}{2} (\bar{y}_1 - \bar{y}_2)' S_{pl}^{-1} (\bar{y}_1 + \bar{y}_2)$$

Linear Discriminant Analysis

- If prior probabilities are known and we assume that the densities are multivariate normal $N_p(\mu_1,\Sigma)$ and $N_p(\mu_2,\Sigma)$ then the classification rule becomes
  - Assign **y** to G_1 if

$$\mathbf{a}'\mathbf{y} = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)'\mathbf{S}_{\mathrm{pl}}^{-1}\mathbf{y} > \frac{1}{2}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)'\mathbf{S}_{\mathrm{pl}}^{-1}(\bar{\mathbf{y}}_1 + \bar{\mathbf{y}}_2) + \ln\frac{p2}{p1}$$

  - Assign **y** to G_2 otherwise

Unequal Misclassification Errors

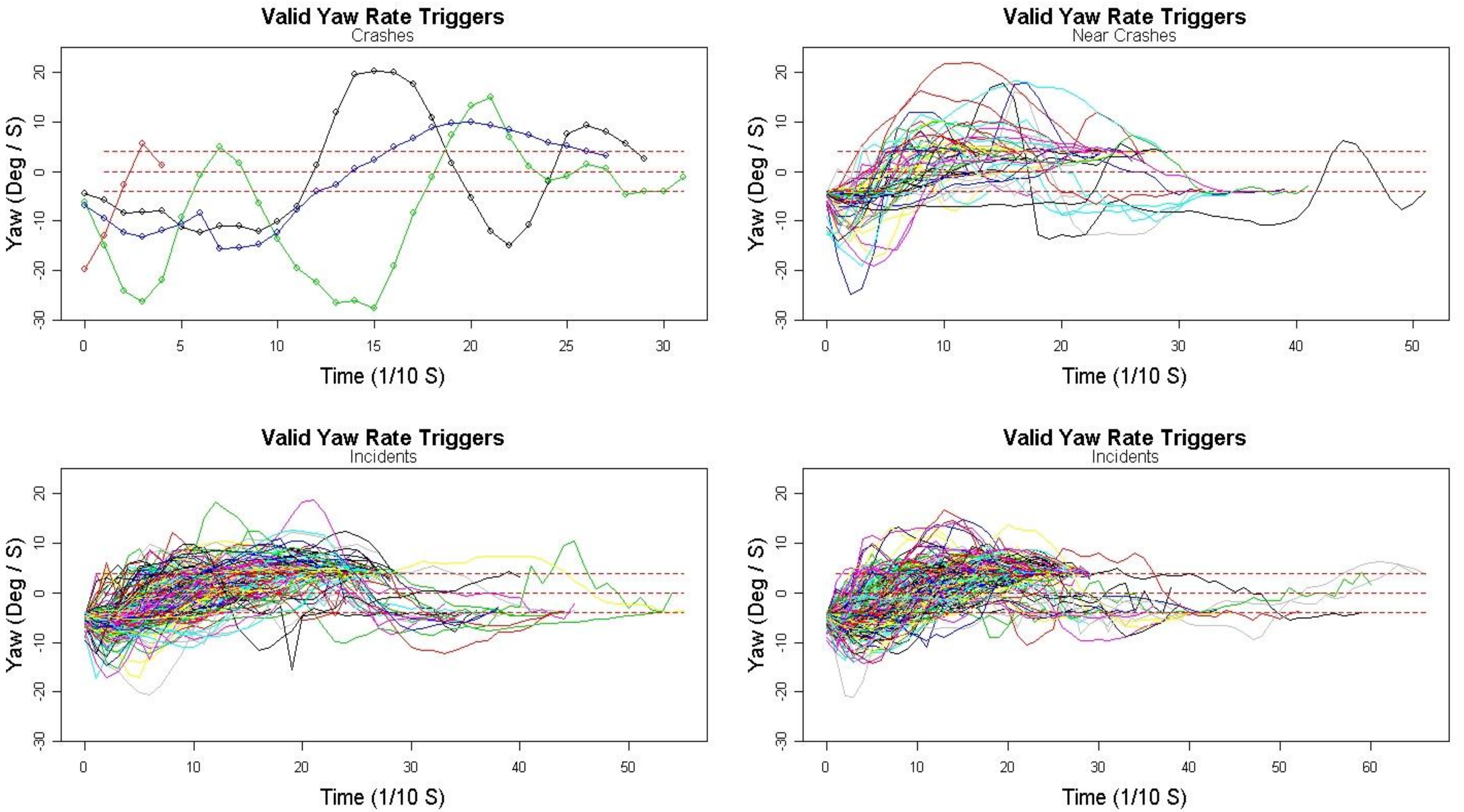
- If misclassification errors differ in severity relative costs can be assigned to the errors and incorporated into classifier
- Let C(i|j) represent the cost of misclassifying an observation from G_j into G_i
- Assign y to G1 if $d_1^* < d_2^*$ where

$$d_i^* = \frac{1}{2}(y - \bar{y}_i)' \Sigma_{pl}^{-1} (y - \bar{y}_i) - \ln\left[p_i C(j \mid i)\right] \text{ for } i \neq j = 1,2$$

Curve Classification

- Another approach to the yaw rate trigger classification problem is a functional formulation
- In this approach interest is focused on the smooth underlying function rather than vectors of observations in discrete time
- The readings for each observation are replaced with a continuous function obtained via basis expansion of the data

Yaw Rate Profile Plot



Curve Classification

- Multivariate data
  - Number of replications >> number of variables
  - No obvious reason or advantage to model the variables as values of a random function
- Functional data
  - Number of replications $\not\gg$ number of variables
  - Achieving dimension reduction by modeling the variables as values of smooth random function
- Functional (high dimensional) approaches
  - Active area of research

Conclusions

- There are aspects of naturalistic data that make classification efforts challenging
- There are a wide assortment of classification methods that may be amenable to naturalistic data classification tasks

References

- Eubank, R., Hsing, T. *Functional Data Analysis*, 32nd Annual Summer Institute of Applied Statistics, Brigham Young University, June 20-22, 2007.
- Holmstrom, L., Koistinen, J., Laaksonen, J. (1997)Neural and Statistical Classifiers-Taxonomy and Two Case Studies, IEEE Transactions on Neural Networks, 8:1.
- Johnson, D.E. Applied Multivariate Methods for Data Analysts, Duxbury Press.
- Ramsay, J., Silverman, B. Applied Functional Data Analysis: Methods and Case Studies, Springer.
- Ramsay, J., Silverman, B. *Functional Data Analysis: Second Edition*. Springer.
- Rencher, A.C., *Methods of Multivariate Analysis*, Wiley Inter-Science New York 1995.

Driving Transportation with Technology