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

University of Michigan
Transportation Research Institute

Aim of project

Capture common elements in how highway factors are associated with

- Crashes (as recorded in crash data)
- Driving behaviors (as recorded by FOT data)

Background

- ***Control*** is defined here as the effectiveness of tactical and operational aspects of the driving task
- Crashes occur when there is a loss of control (for whatever reason)
- Not every loss of control event results in a crash
- Road departure crashes
- FOT data – contain indicators of control problems
  - Certain lane departure warning
  - Others - to be determined

Hypotheses

- Crash surrogates exist and can be found in FOT data
- The crash surrogates are related to highway features and to driver-related factors
- Crash surrogates are related to control

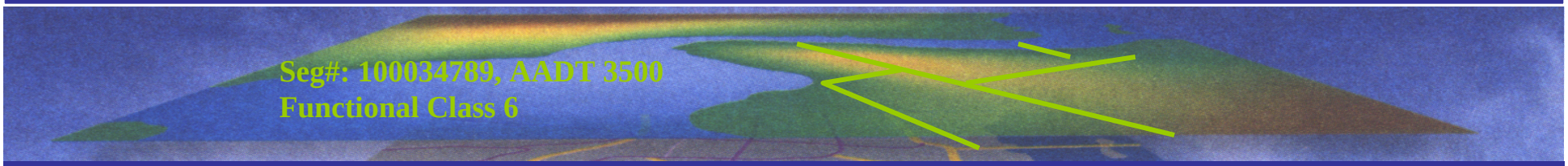
Data

- 71,000 relevant crash events in SE Michigan from 2001-2005 crash data
- 220,000 miles FOT data in SE Michigan
- 8,300 road departure warnings (~0.1 per vmt)
- Roadway information
  - Michigan Base GIS
  - HPMS

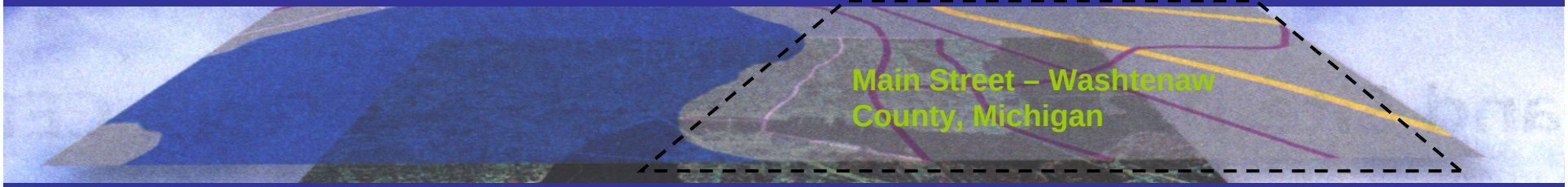
Currently 4 sources of data are spatially referenced either using existing GPS data or completing spatial joins using GIS software utilities.

Layer One - Michigan Run-Off-Road Crashes-Spatial Data (Lat/Long)

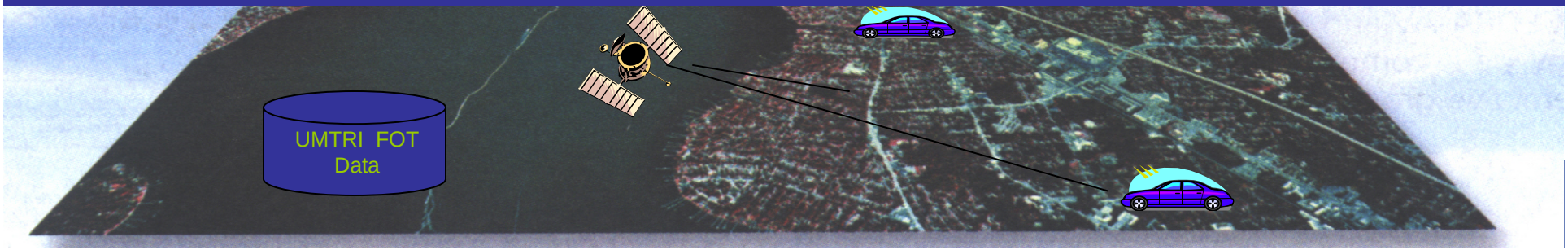
Layer Two - Michigan 2005 HPMS- Table Data



Layer Three - State of Michigan public roads base map

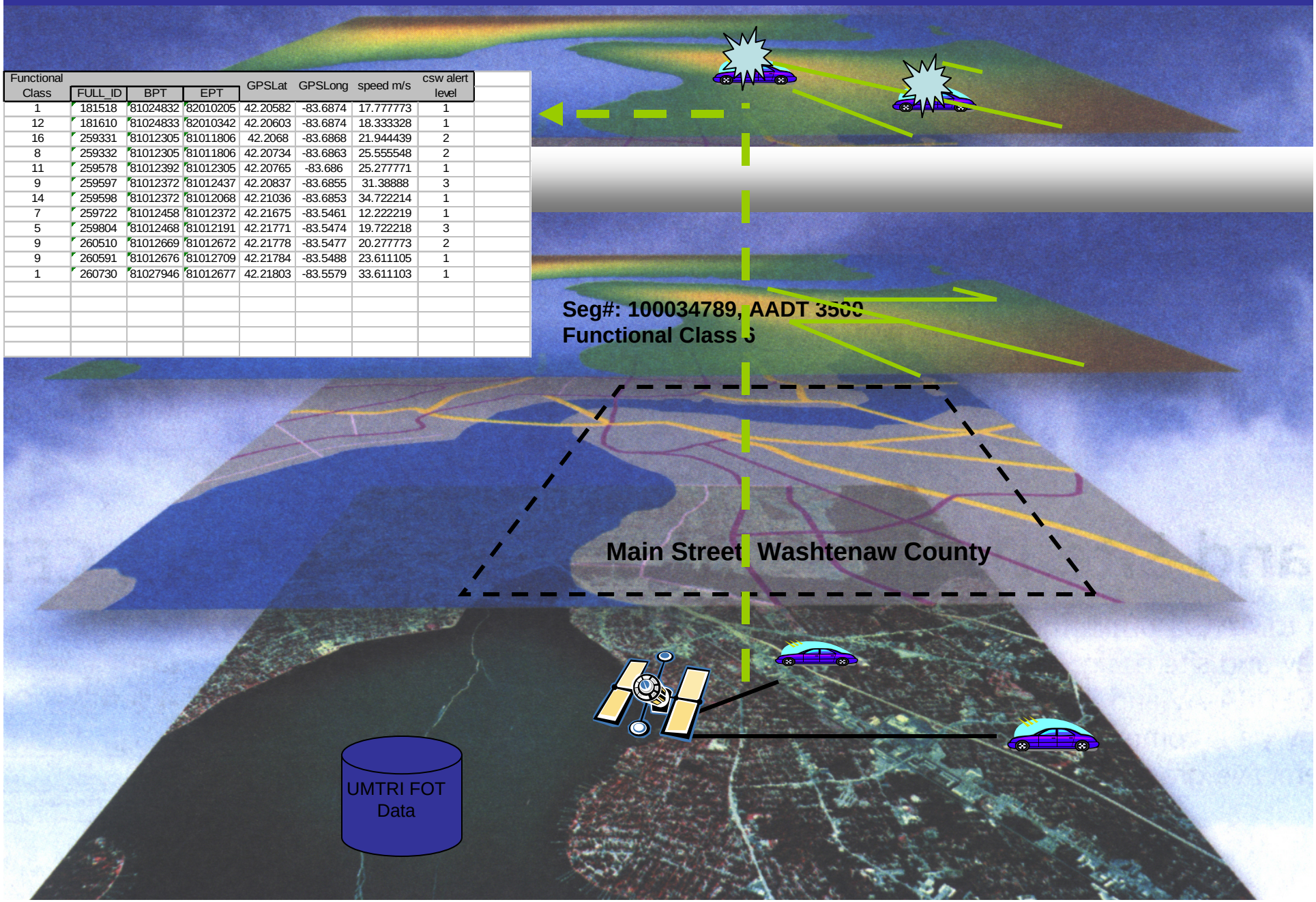


Layer Four - UMTRI Naturalistic driving data from FOT fleets including Lat Long Positions



These sources are “joined” to create a flat file for analysis of research questions.

Functional Class	FULL_ID	BPT	EPT	GPSLat	GPSLong	speed m/s	csw alert level
1	181518	81024832	82010205	42.20582	-83.6874	17.777773	1
12	181610	81024833	82010342	42.20603	-83.6874	18.333328	1
16	259331	81012305	81011806	42.2068	-83.6868	21.944439	2
8	259332	81012305	81011806	42.20734	-83.6863	25.555548	2
11	259578	81012392	81012305	42.20765	-83.686	25.277771	1
9	259597	81012372	81012437	42.20837	-83.6855	31.38888	3
14	259598	81012372	81012068	42.21036	-83.6853	34.722214	1
7	259722	81012458	81012372	42.21675	-83.5461	12.222219	1
5	259804	81012468	81012191	42.21771	-83.5474	19.722218	3
9	260510	81012669	81012672	42.21778	-83.5477	20.277773	2
9	260591	81012676	81012709	42.21784	-83.5488	23.611105	1
1	260730	81027946	81012677	42.21803	-83.5579	33.611103	1



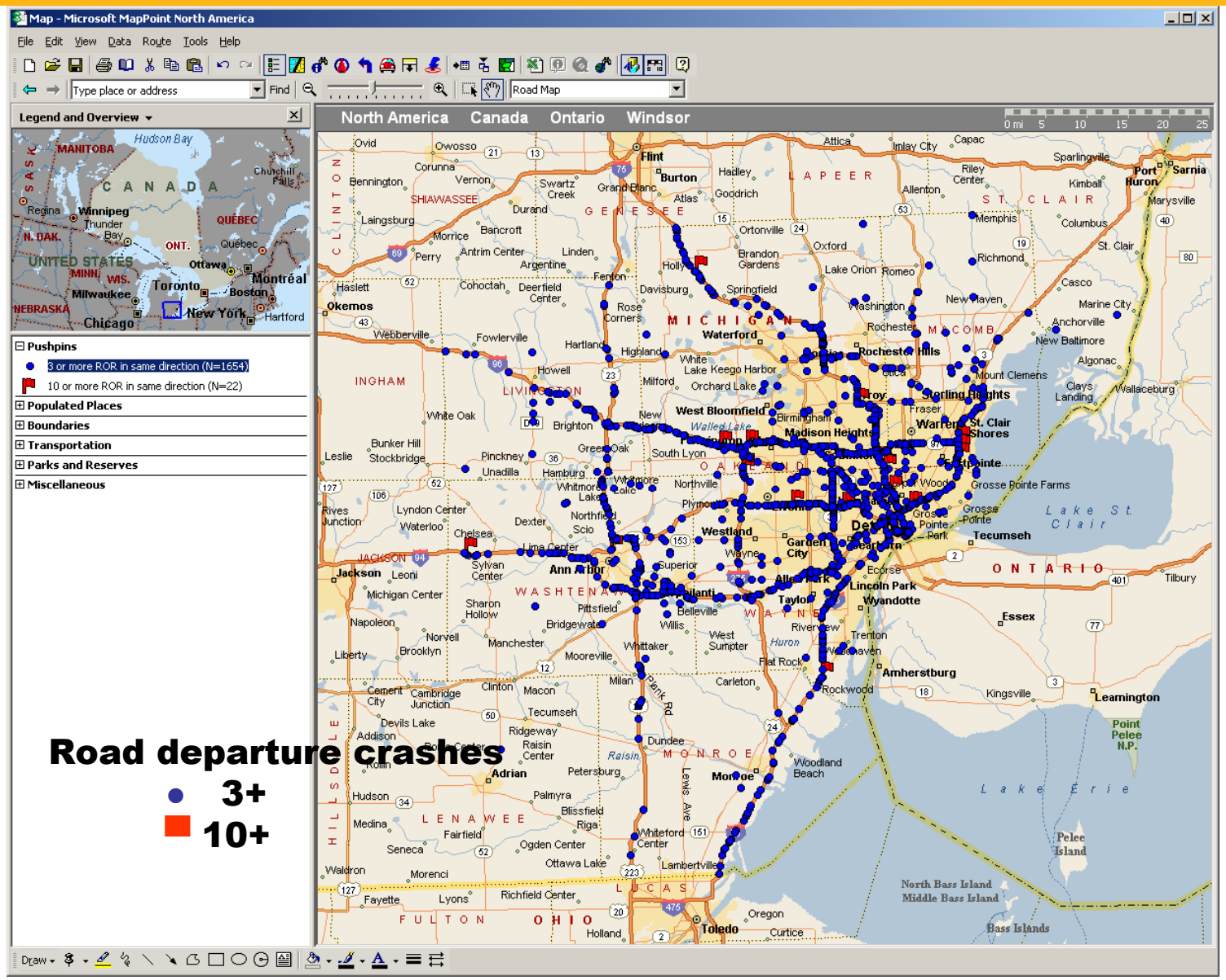
Additional sources of spatial data being collecting from selected agencies include:

- -intersections
- -signalized intersections
- -traffic control signage locations
- -aerial photographs-pavement condition data

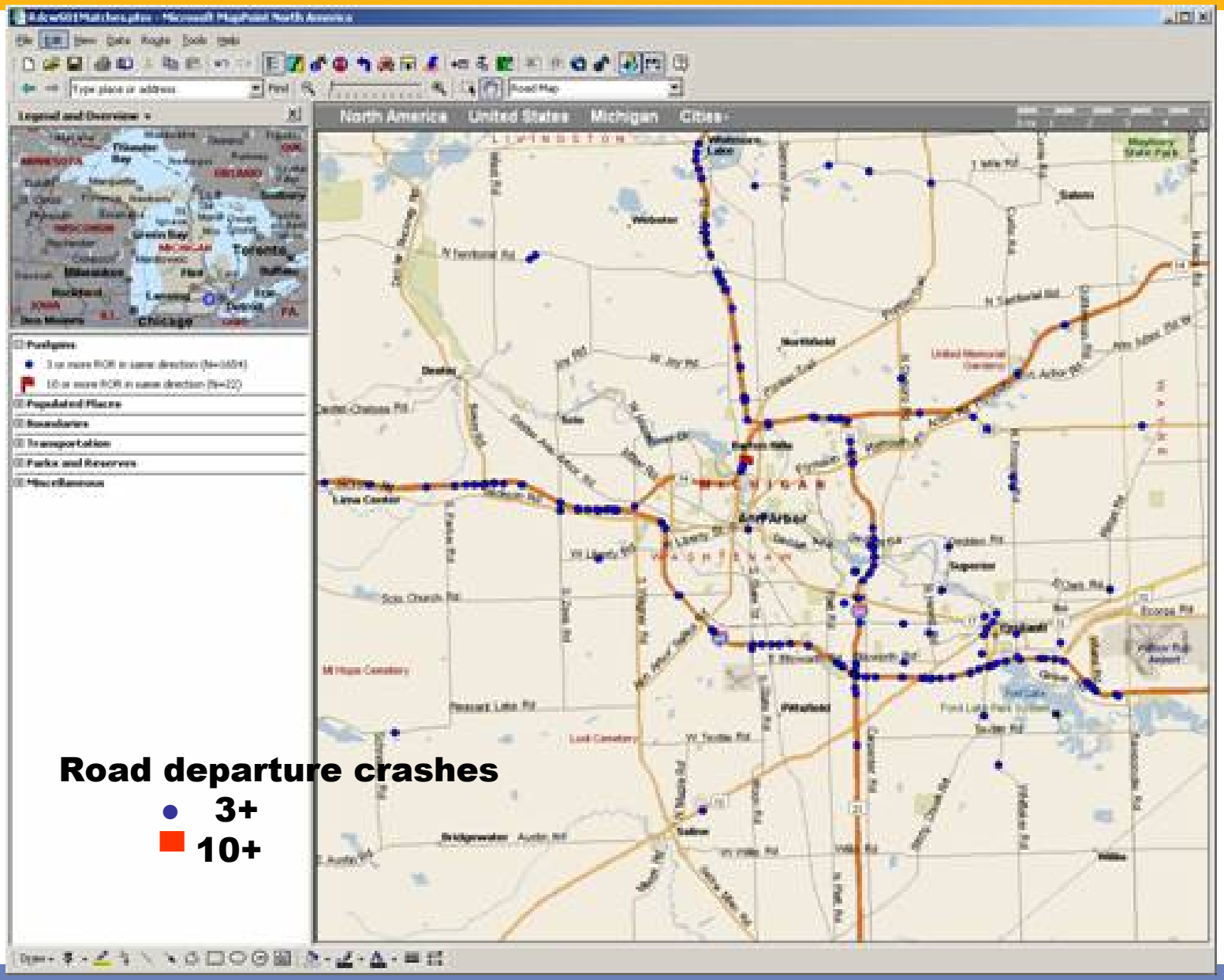
UMTRI

Overlap of FOT driving locations and road departure crashes

- 37,117 locations+ direction where FOT driving data overlap road departure crash
- 22 locations with 10+ road departure crashes within 20 meters of each other
- 1,654 locations with 3+ road departure crashes within 20 meters of each other


Map - Microsoft MapPoint North America
File Edit View Data Route Tools Help
Type place or address
Find
Road Map
Legend and Overview
Pushpins
8 or more ROR in same direction (N=1654)
10 or more ROR in same direction (N=22)
Populated Places
Boundaries
Transportation
Parks and Reserves
Miscellaneous
North America Canada Ontario Windsor
0 mi 5 10 15 20 25
Road departure crashes
3+
10+
Draw

UMTRI

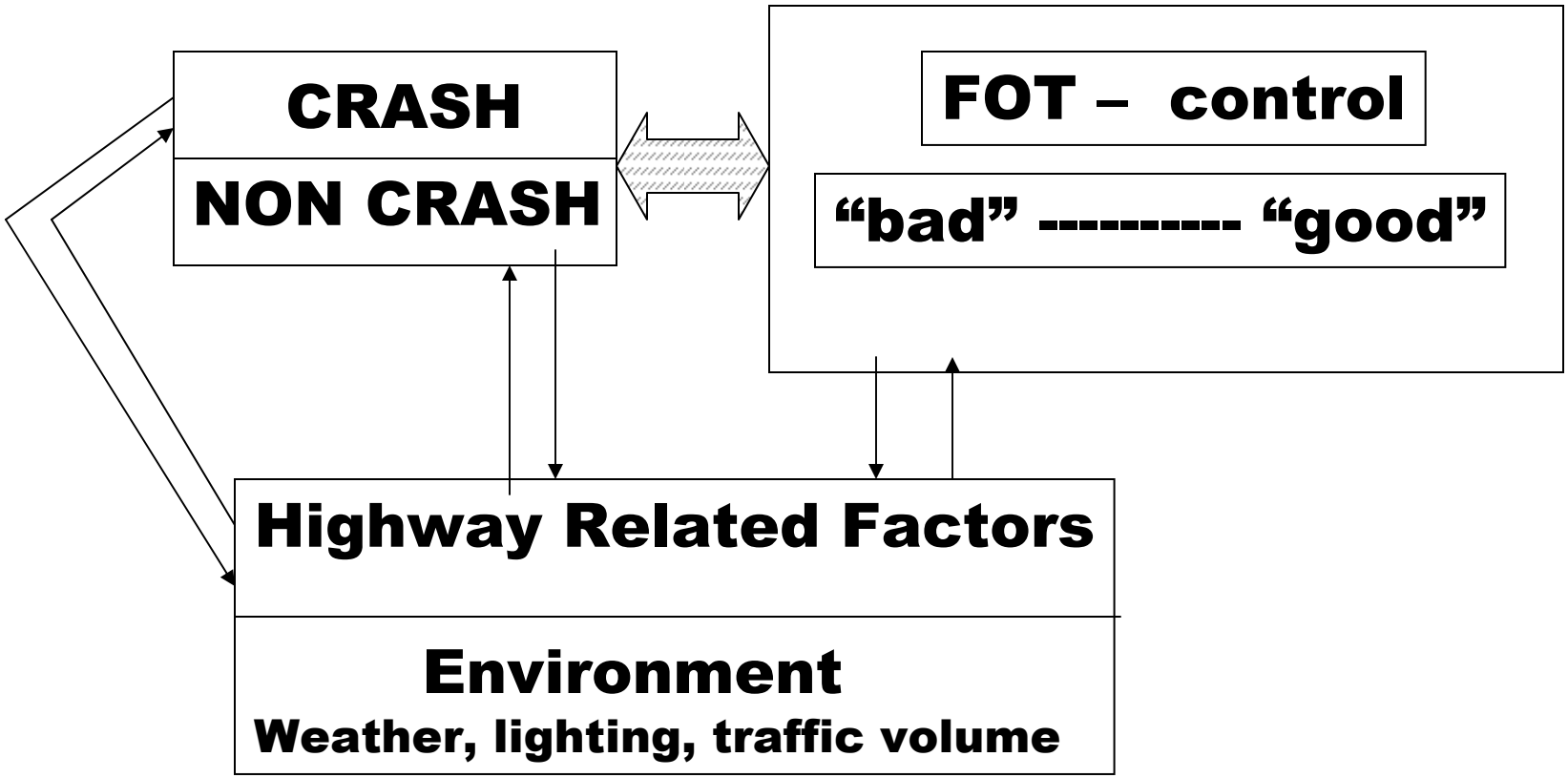

North America
United States
Michigan
Cities
Road departure crashes
3+
10+

UMTRI

Sites with FOT data
and at least
3 road departure crashes

UMTRI

Experimental Design



Crash/Non Crash events

- Non crash
  - Road Segment with no crashes
    - VMT, geometric features

Examine crash/non crash events

- Literature review
- Clustering of crashes at locations
- Patterns in highway factors
- What is relevant in road departure crash

Control metrics

- Identify discriminators of “good” “bad” control
- What are the patterns of control indicators at crash and non crash locations?
  - Discrete, Continuous, dynamic sequence?
- Can the patterns of control indicators be used to predict high crash locations?
- Do they satisfy surrogate evaluation criteria?

Driver Factor Questions

- Can relevant driver factors be identified?
  - Aggressive driving
  - Distracted driving
  - Fatigued driving
  - Engaged driving
- Can their influence be controlled for in analysis?

Challenges for Year 1

- Refine definitions and hypotheses
- Experimental design
- Tests of reasonableness
- Supporting statistical models and analyses methods

Crash Data and Models

Response	Crashes
	(Exposure)
Explanatory Factors	Driver
	Environment
	Vehicle
	Highway

Standard models are available for assessing risk in crash data

Example: Hierarchical Bayesian Models

Likelihood: Poisson
First Prior: Gamma or Normal (attach mean to log-linear model with explanatory variables)
Second Prior: noninformative proper priors on model parameters

Crashes and Surrogate Measures

But this project involves an additional component.

Can surrogates be related to actual crashes?

Crash Data

Response	Crashes
	(Exposure)
Explanatory Factors	Driver
	Environment
	Vehicle
	Highway

FOT Data

Response	Surrogates
	(Exposure)
Explanatory Factors	Driver
	Environment
	Vehicle
	Highway

A Multivariate Generalized Model

$$r_{i1}, r_{i2} \mid \beta_1, \beta_2, \Sigma_i \sim N_2\left(\begin{bmatrix} x_{i1}^T\beta_1 \\ x_{i2}^T\beta_2 \end{bmatrix}, \Sigma_i\right) \qquad i = 1, \ldots\ldots, N$$

log crash rate: $r_{i1} = x_{i1}^T\beta_1 + \varepsilon_{i1} = \beta_{10} + \beta_{11}x_{i11} + \ldots + \beta_{1p}x_{i1p} + \varepsilon_{i1}$

log surrogate rate: $r_{i2} = x_{i2}^T\beta_2 + \varepsilon_{i2} = \beta_{20} + \beta_{21}x_{i21} + \ldots + \beta_{2q}x_{i2q} + \varepsilon_{i2}$

Relative risks → $\beta = [\beta_1 \mid \beta_2]$

crashes **surrogates**

$\mathrm{Cov}(\varepsilon_{i1}, \varepsilon_{i2}) = \Sigma_i$

A variance stabilizing transformation is available

$H_0 : \beta_1 = \beta_2$ **excluding intercepts** β_{10}, β_{20}

$H_0 : \beta_{11} = \beta_{21}$

UMTRI

Advantages of the Multivariate Generalized Model

- WLS can be used to estimate ML parameters in Poisson log-linear models. The proposed model is a multivariate generalization of WLS. (The marginal models for crashes and surrogates are equivalent to the usual univariate WLS models).
- The model incorporates a correlation structure between crashes and crash surrogates (through Σ).
- Overdispersion relative to the Poisson model is handled through the scale parameters in the Normal distribution.
- All calculations are tractable and fast. Models used in econometrics (no need for MCMC simulation).
- Formal hypothesis tests can be carried out to determine if surrogates are good measures of crash risk.

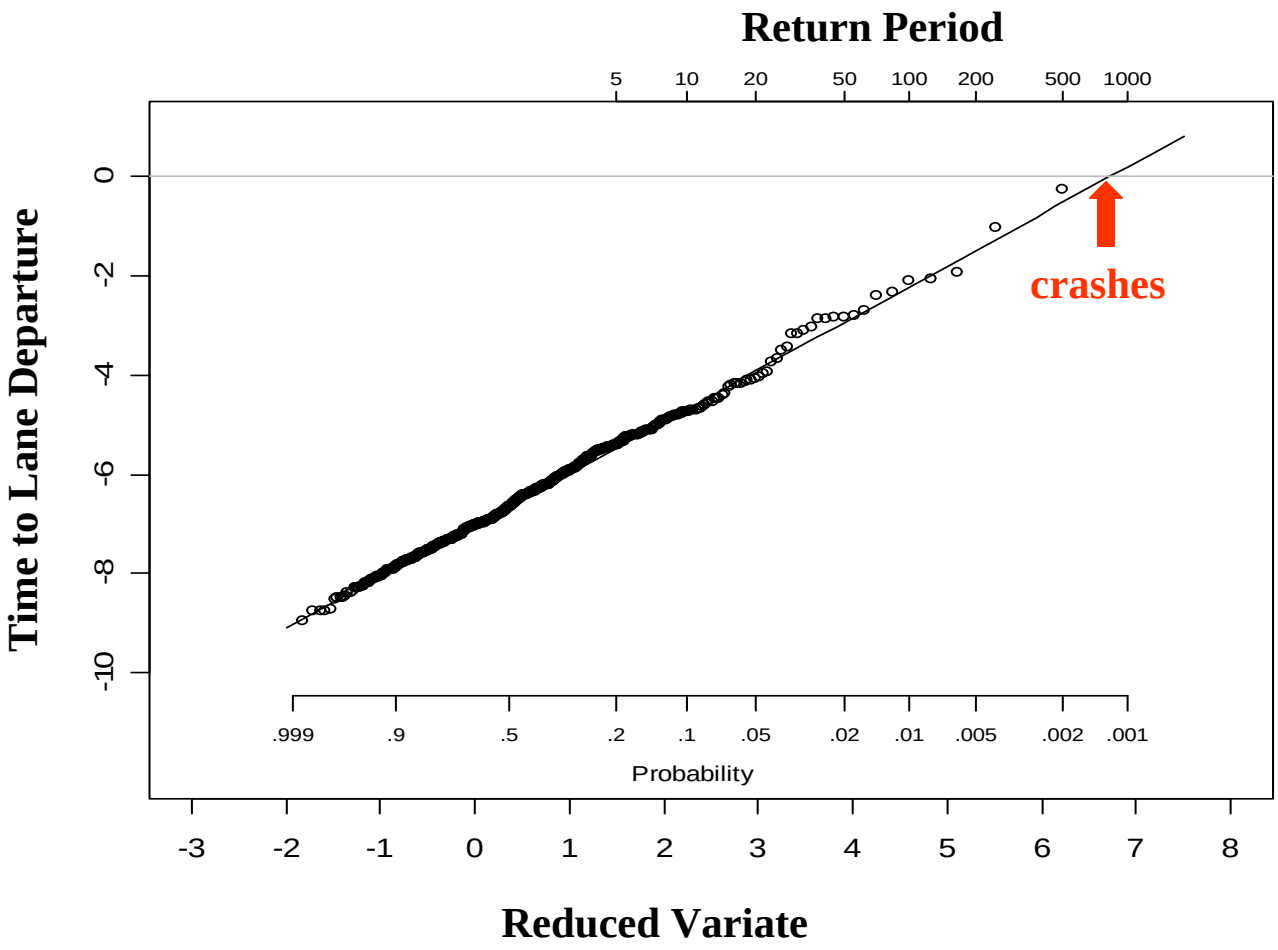
Extreme Value Theory

- Idea is to model rare events that lie outside the range of available observations.
- The smallest observation is selected from each of many samples.
- The resulting sample of minimum values is the sample of extreme values.

Example: In each time frame of 15 min, record the minimum time to road departure.

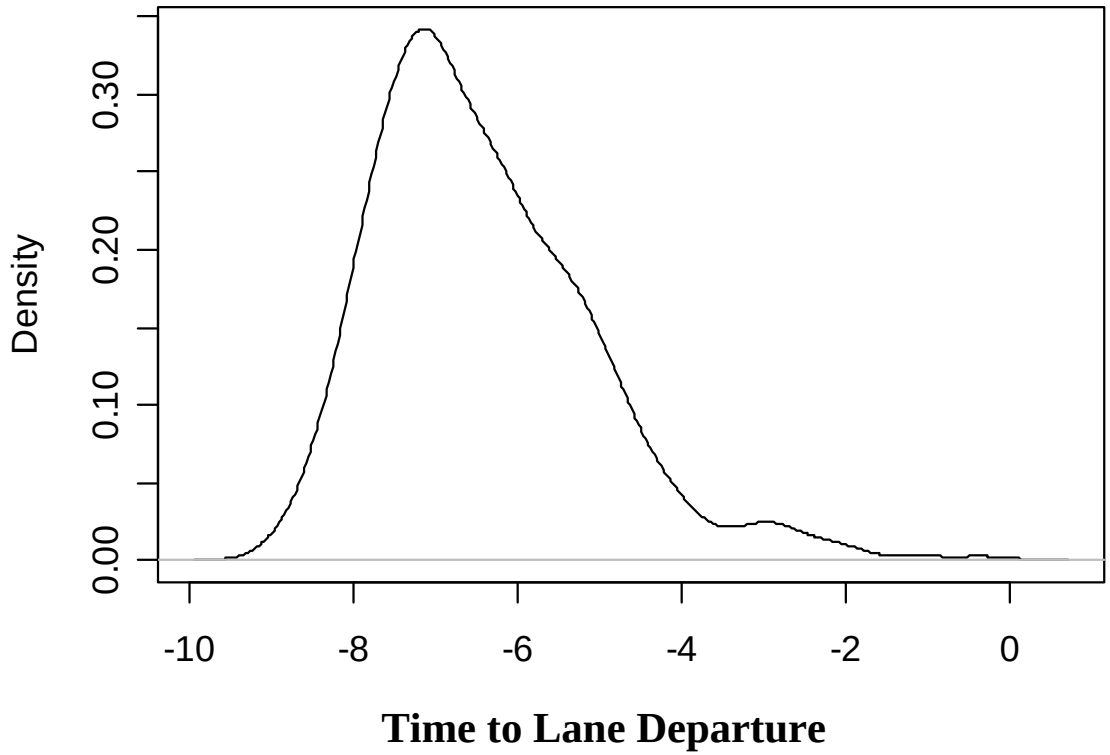
- EVT can also be approached in a Bayesian framework in a regression setting.

Extreme Value Theory



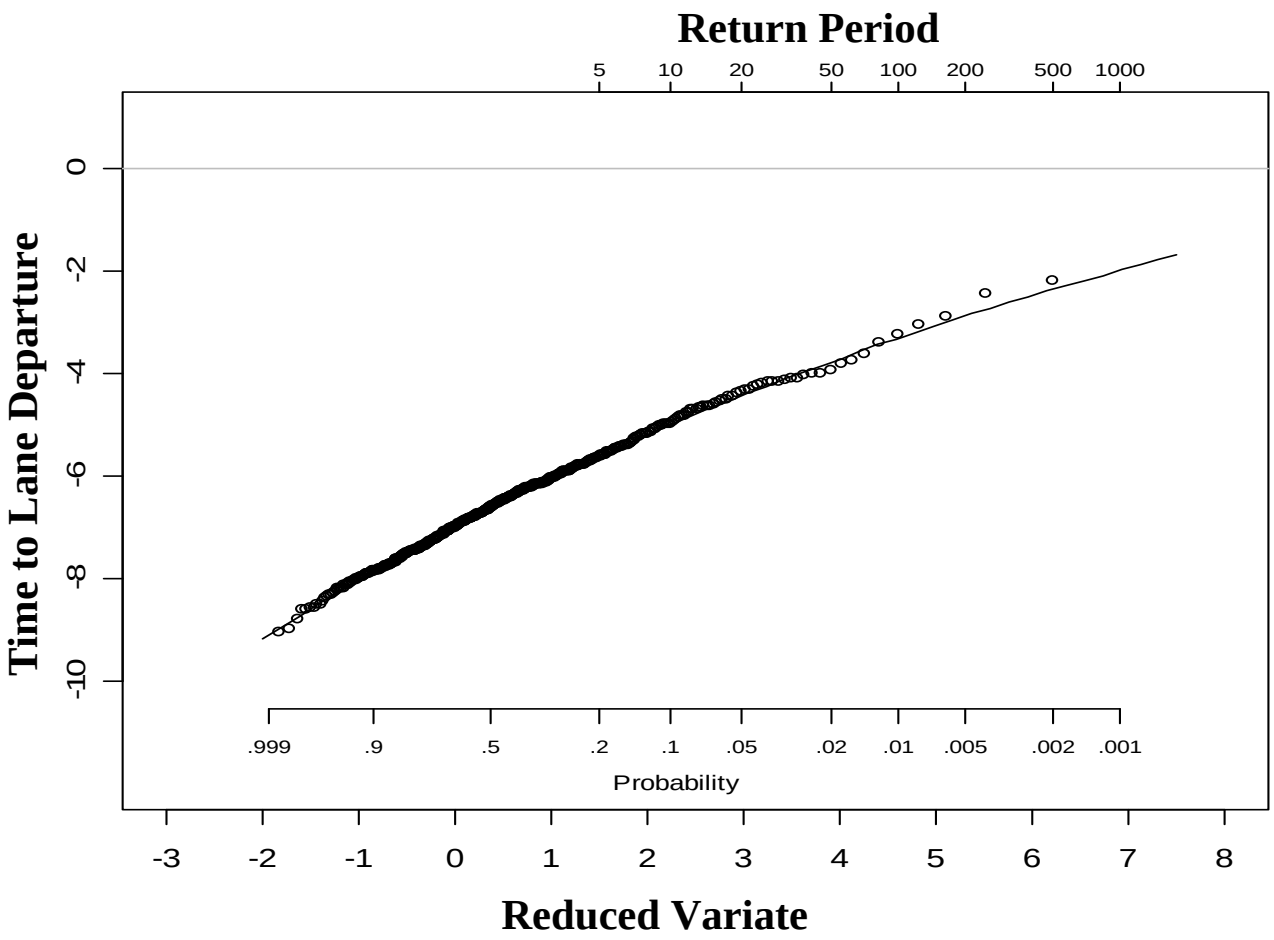
UMTRI

Extreme Value Theory



UMTRI

Extreme Value Theory



UMTRI

Thank You !