

Extreme Value Theory for Measuring Safety

Andrew P. Tarko
Center for the Advancement of Transportation Safety
School of Civil Engineering
Purdue University

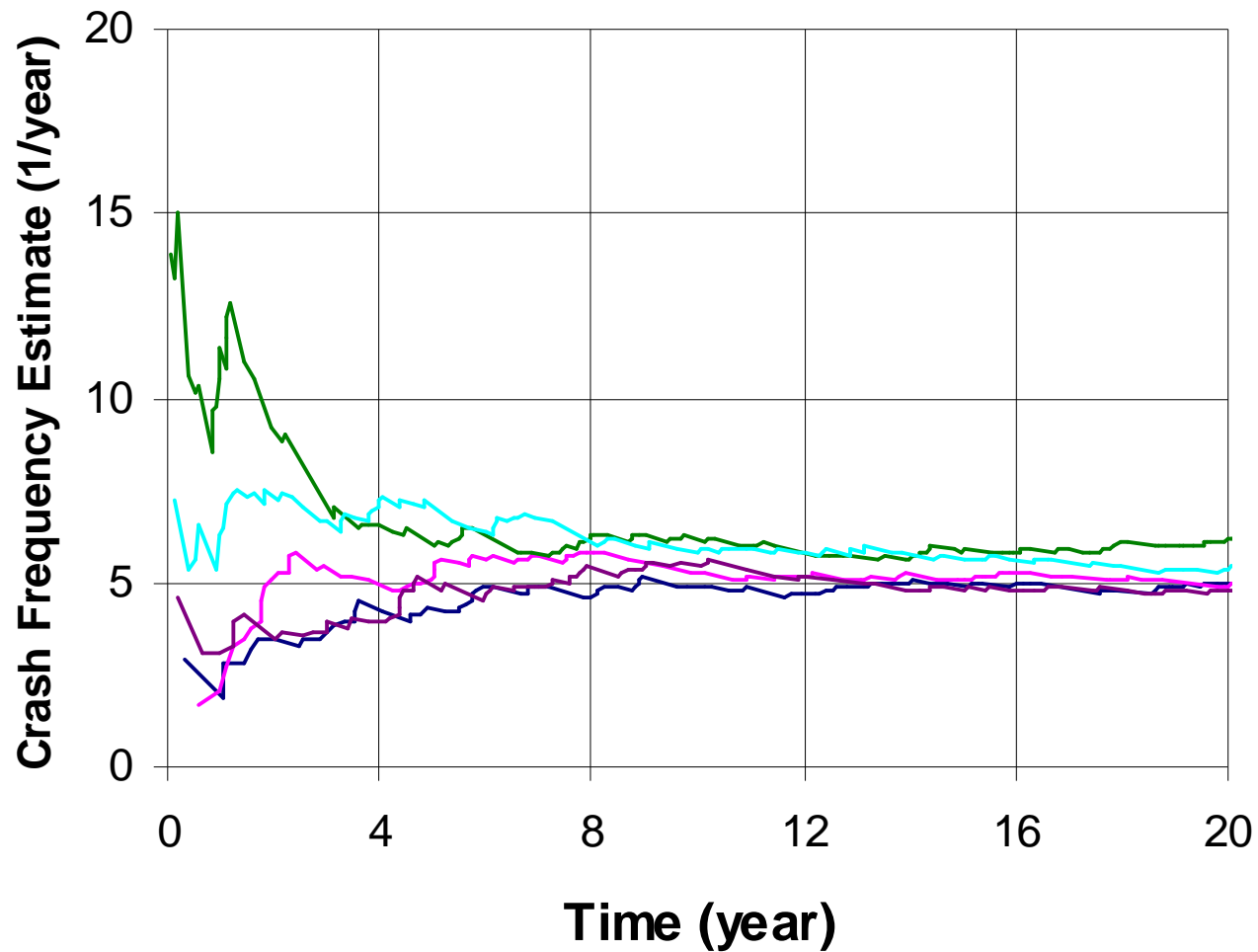
SHRP II Safety Research Symposium
Washington, D.C.
August 16-17, 2006

Measuring Highway Safety

- Safety research and management require measuring safety
- Desirable measure - crash frequency and severity
- Direct measurement – counting reported crashes
- Crash counts is a Poisson-like process

Measurement Efficiency

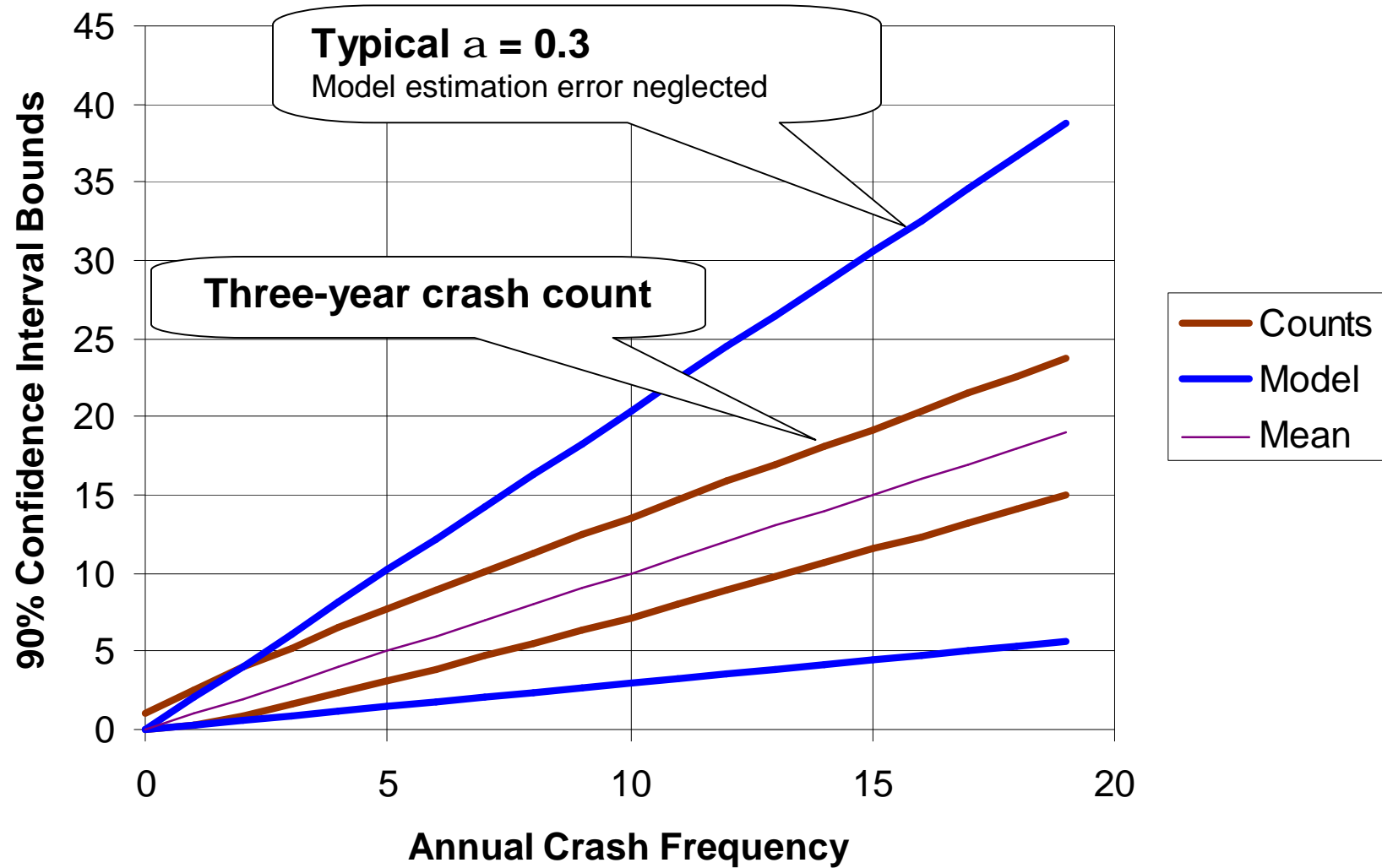
Crash Counts



Improving Measurement Efficiency

- Regression models
- Combining regression estimates with crash counts (EB method)

Measurement Efficiency of Regression



Traffic Conflict Technique for Highways

- Traffic Conflicts Technique
 - Traffic conflict between vehicles if at least one of the drivers has to make rapid evasive maneuver to avoid collision
- Fundamental assumption:

$$\textit{Crash Frequency} = \mathbf{k} \cdot \textit{Conflict Frequency}$$

- Fundamental assumption unconfirmed
- Proportionality factor undetermined

Traffic Conflict Technique

Statistical Consideration

- Why have over 30 years of research not delivered a convincing statistical evidence of existence of the crash-conflict relationship?
- Possible explanations:
 1. Measurement errors
 - Poisson variability of crash and conflict counts
 - Deficient crash data
 - Vague definition of traffic conflict and measurement difficulties
 2. The proportionality factor k varies across locations

Crash Occurrence as a Binomial Process

$$a = E \cdot R$$

where:

a = annual crash frequency or expected annual C ,

E = annual exposure (traffic volume, miles traveled),

R = average probability of crash (risk) under the unit exposure.

$$\text{var } C = E \cdot R \cdot (1-R) \stackrel{R \ll 1}{=} E \cdot R = a$$

Existing Safety Measurement Methods

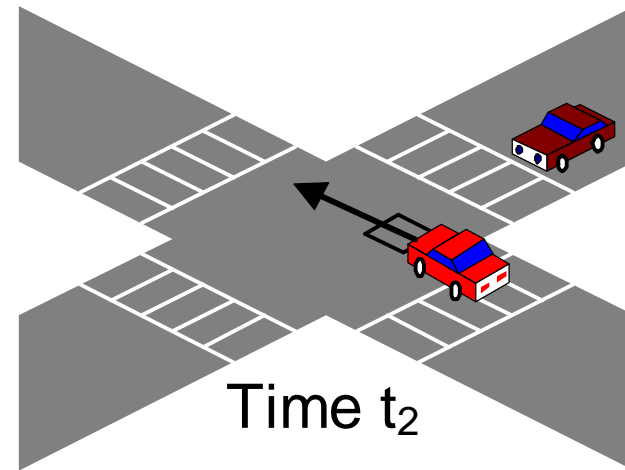
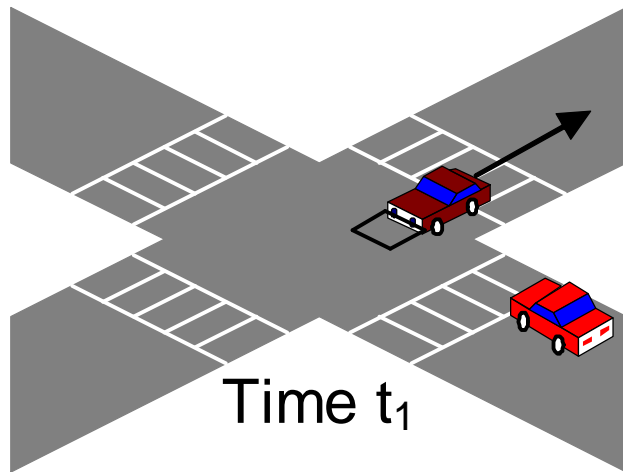
Method	Exposure E	Risk R	Crash frequency a
Crash counts c	Time t	NA	c/t
Regression	Volume, VMT	f(road traits)	$E \cdot f(\text{road traits})$
Bayesian method	NA	NA	$w_1 \cdot (c/t) + w_2 \cdot E \cdot f(\text{road traits})$
Traffic conflicts N	Frequency of conflicts N	Constant k	$k \cdot N$

Proposed Approach

1. Observe dangerous events (similar to traffic conflicts)
2. Measure crash proximity of these events
3. **Crash proximity must have a definitive boundary between conflict and crash**
4. **Estimate likelihood of crash (R) as an extreme event**
5. Estimate frequency of hazardous events (E)
6. Estimate crash frequency $a = E \cdot R$

Proposed Approach

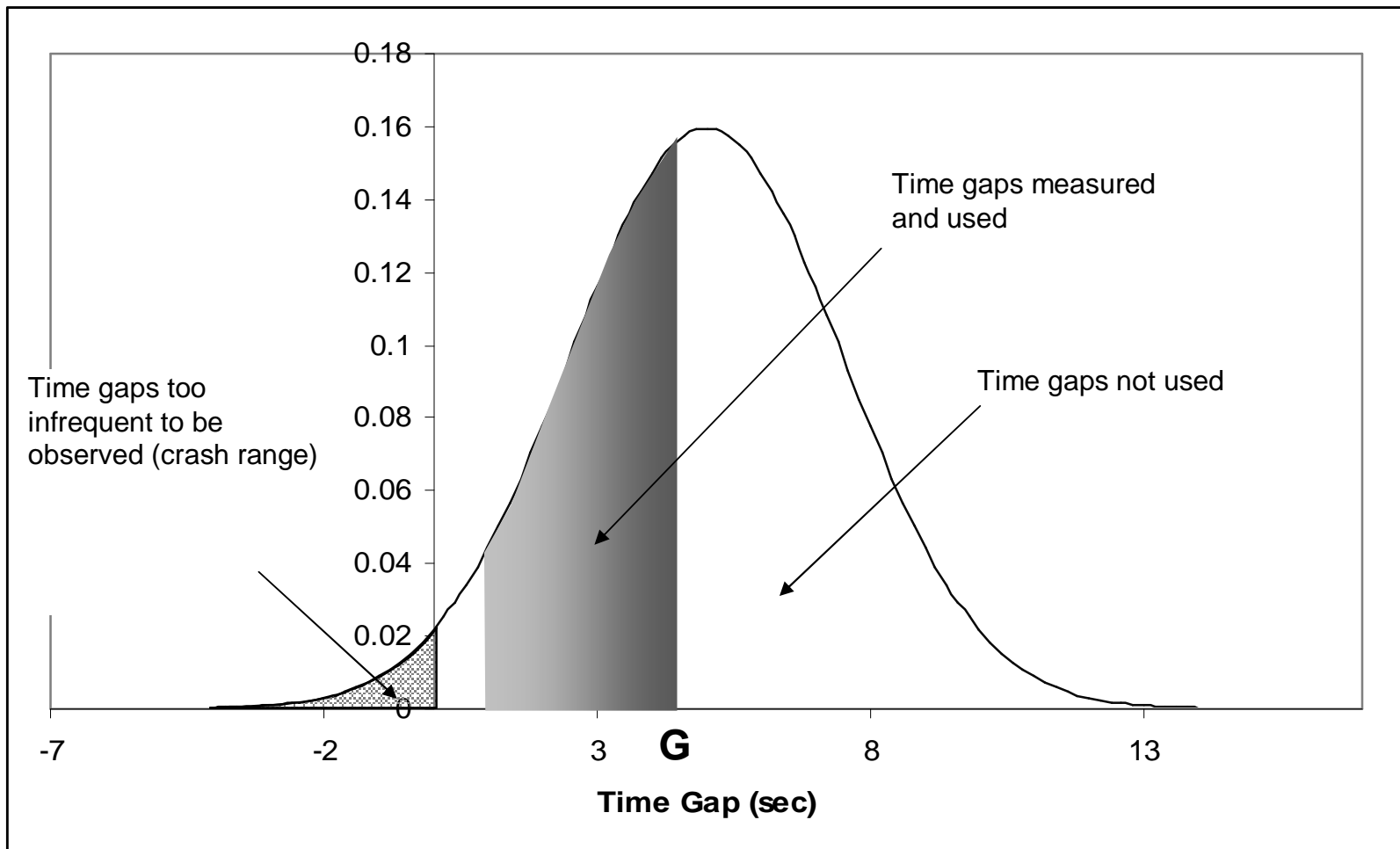
Dangerous Event



Dangerous event if $g = t_2 - t_1 < G$

Proposed Approach

Crash Proximity g



Example EVT Model

Pareto Distribution

$$z = G - g$$

$$F(z) = 1 - \left(1 + \frac{x \cdot z}{s}\right)^{-1/x}, \text{ for } s > 0.$$

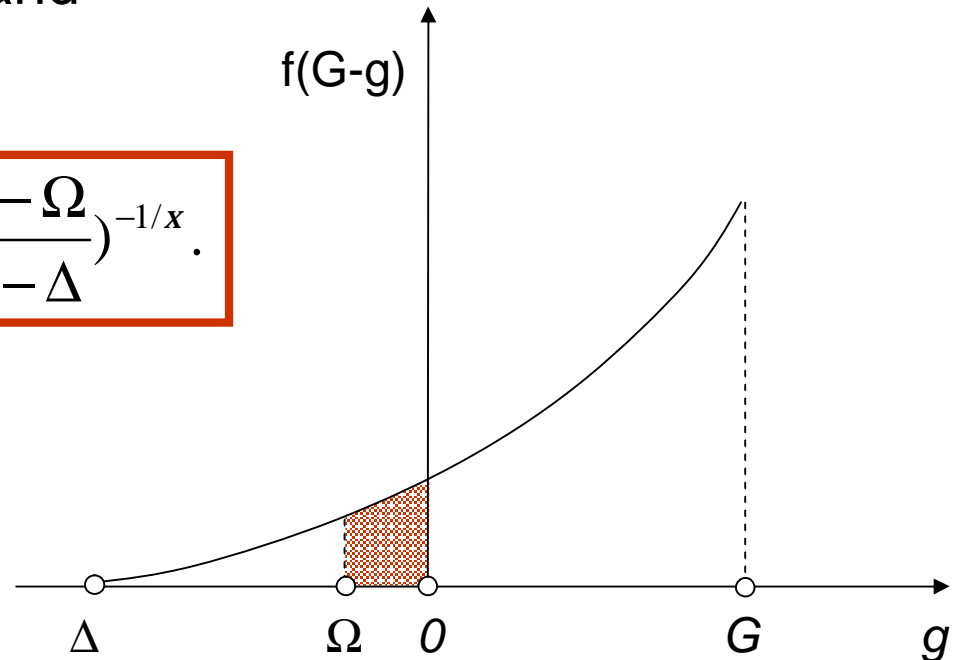
If Δ is the lowest value: $F(z = G - \Delta) = 0$, then

$$F(g) = 1 - \left(1 - \frac{G - g}{G - \Delta}\right)^{-1/x} \text{ and}$$

$$R = F(0) - F(\Omega)$$

$$R = \left(1 - \frac{G}{G - \Delta}\right)^{-1/x} - \left(1 - \frac{G - \Omega}{G - \Delta}\right)^{-1/x}.$$

where $-\Omega$ is the time a vehicle spends in the collision zone



Example EVT Model

Estimating Pareto Distribution

Let g_i ($i=1..n$) be n realizations of a Pareto variable.
The log-likelihood function is

$$\mathbf{l}(\mathbf{x}, \Delta) = -n \ln[-\mathbf{x}(G - \Delta)] - \left(1 + \frac{1}{\mathbf{x}}\right) \sum_{i=1}^n \ln\left(\frac{g_i - \Delta}{G - \Delta}\right)$$

and ML estimator of ξ is (Δ known):

$$\mathbf{x} = \frac{1}{n} \sum_{i=1}^n \ln\left(\frac{g_i - \Delta}{G - \Delta}\right)$$

Example EVT Model

Estimating Collision Frequency

$$a = E \cdot R$$

$$a = \left(\frac{n}{t}\right) \cdot \left[\left(1 - \frac{G}{G - \Delta}\right)^{-1/x} - \left(1 - \frac{G - \Omega}{G - \Delta}\right)^{-1/x} \right]$$

where:

G = the threshold gap,

ξ and Δ = Pareto distribution parameters,

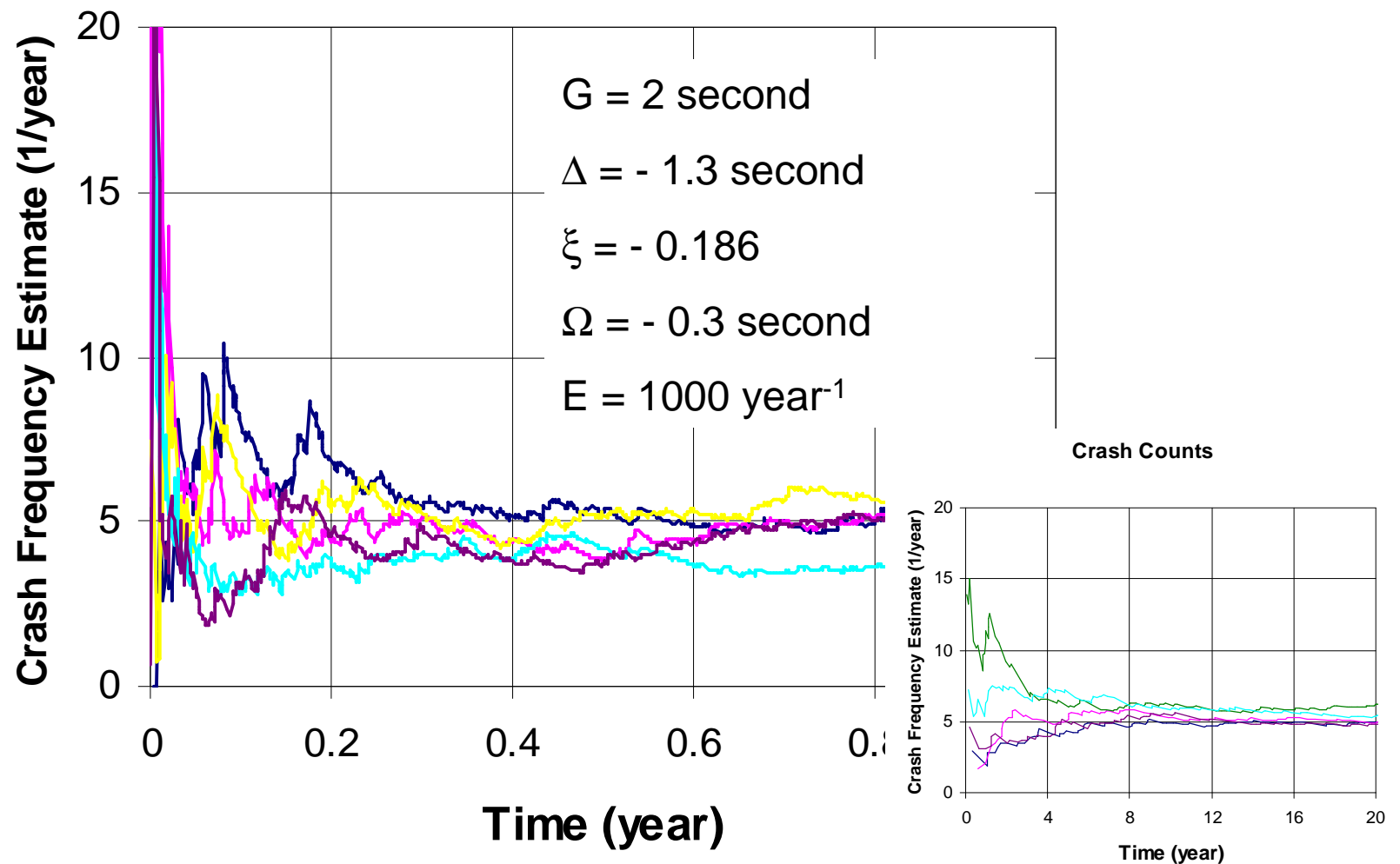
$-\Omega$ = average time spent in the collision zone,

n = number of gaps shorter than G ,

t = observation time.

Example EVT Model

Estimation Efficiency Crossing Time Gaps

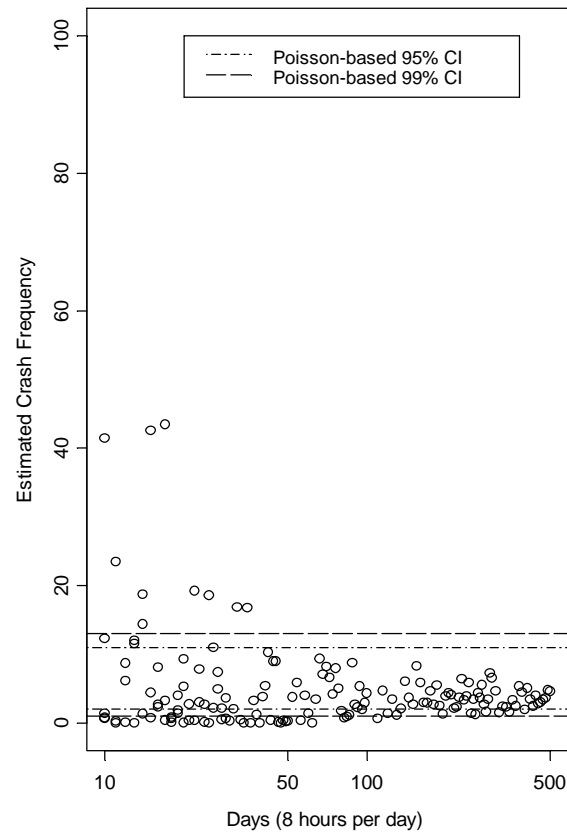


Other EVT Models

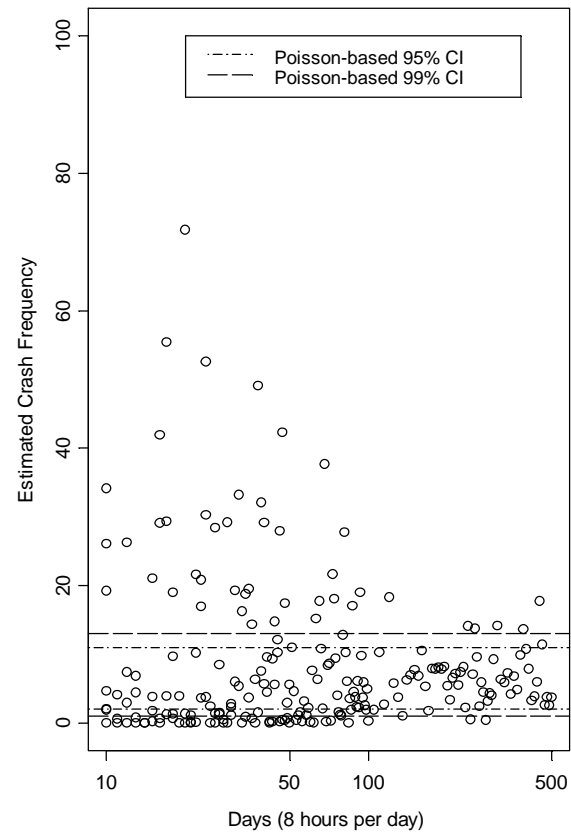
Generalized Extreme Value (GEV) model (Songchitruksa and Tarko, 2006)

$$G(z) = \exp \left\{ - \left[1 + x \left(\frac{z - m}{s} \right) \right]^{-1/x} \right\}$$

Site 87906, r=1 (GEV), Block=15 min



Site 87906, r=1 (GEV), Block=1 hour



Other EVT Models

r-largest Order Statistic Model

$$G_r(z) = \exp(-t(z)) \sum_{s=1}^{r-1} \frac{t(z)^s}{s!}$$

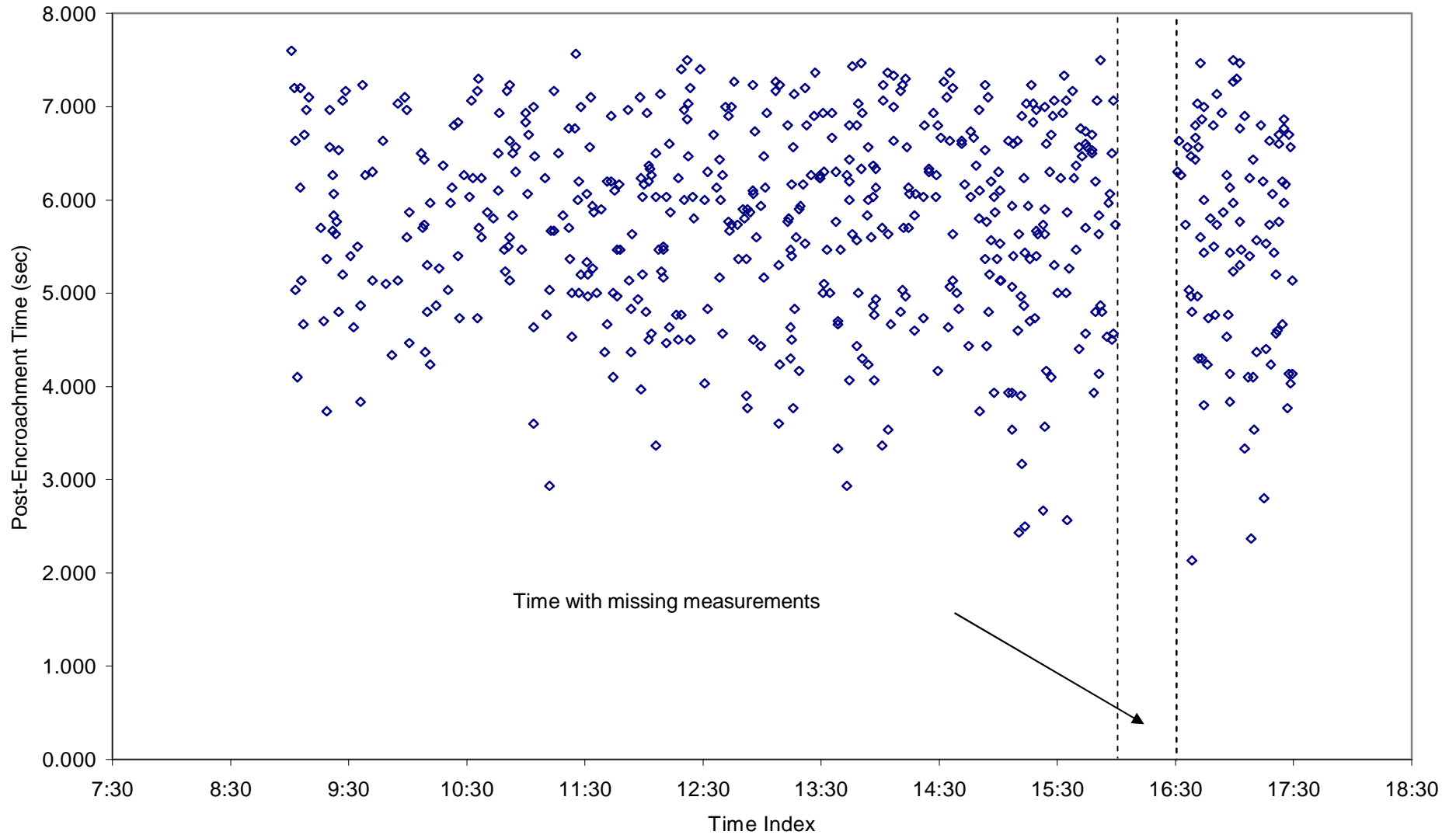
$$t(z) = \left(1 + \mathbf{x} \frac{z - u}{S}\right)^{-1/\mathbf{x}}$$

Further generalization of GEV model

Example Gap Observations

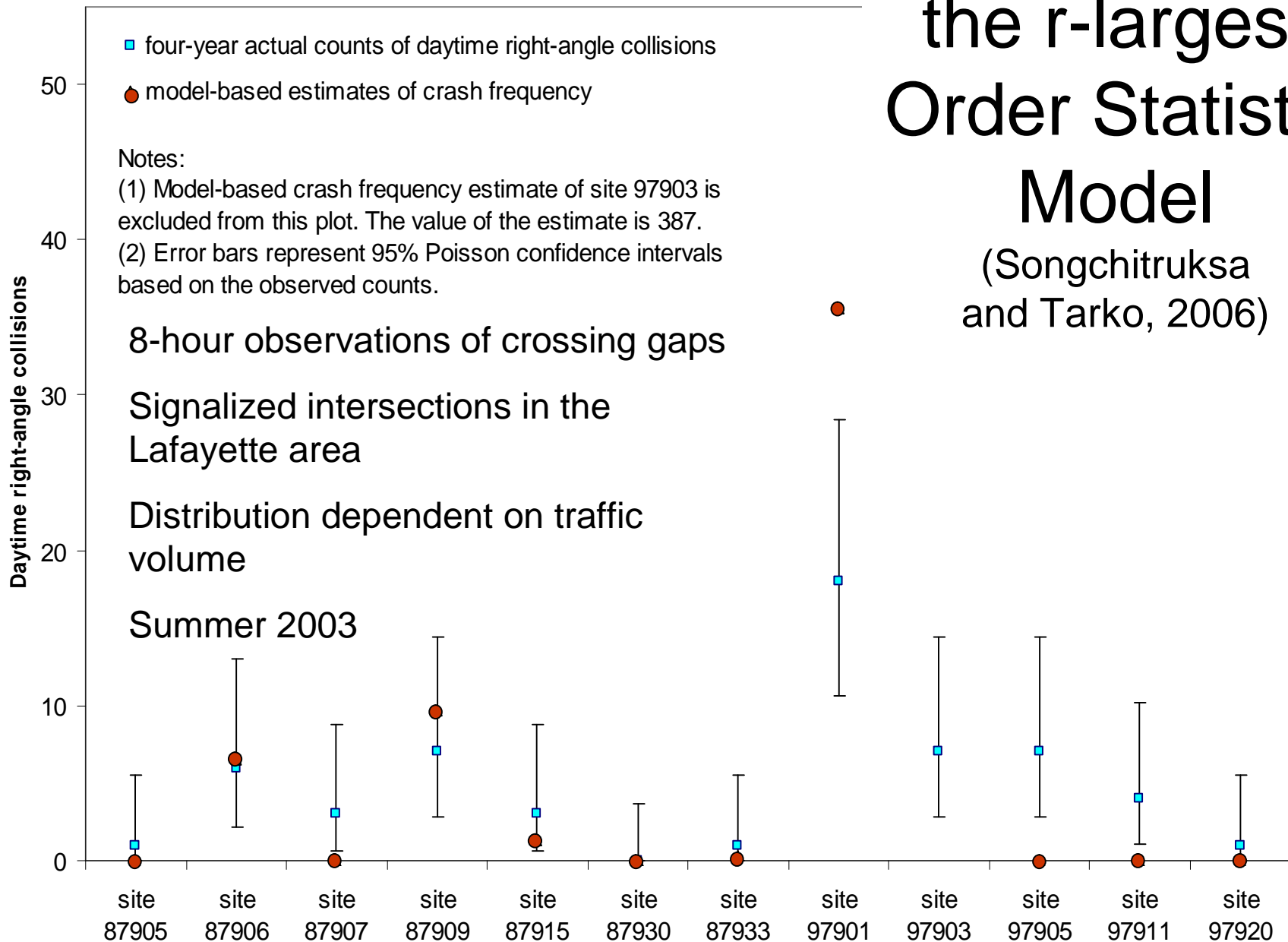
(Songchitruksa and Tarko, 2006)

Intersection: SR-26 @ 18th St, Lafayette, IN (April 8, 2003)



Evaluation of the r-largest Order Statistic Model

(Songchitruksa and Tarko, 2006)



Remarks

- The EVT approach DOES NOT propose any safety surrogates. It estimates CRASH frequency.
- It is based on sound statistical assumptions; meeting these assumptions guarantee validity of the results.
- The key question is the EVT methods EFFICIENCY. It takes 3-5% of the time required by the direct method.

Remarks

- The EVT approach can be used in vehicle-based and site-based studies.
- A good measuring technique is needed to produce input for an EVT method.
- The crash frequency estimate reflects the period of microscopic observations.

Conclusions

- Current safety measurement methods are not able to keep up with the anticipated fast changes in highway safety.
- Current methods cannot be easily used to timely evaluate safety impacts of new solutions.
- Proposed class of EVT methods promises fast results, does not require crash data but returns crash frequency estimates.
- Adequate techniques of measuring traffic characteristics are needed for EVT methods implementation.
- Research is urgently needed to start developing EVT methods for selected applications, gaining knowledge, and improving/expanding the original concepts.