An Assessment of Crime Forecasting Models

FCSM Research and Policy Conference
Washington DC, March 9, 2018

HAUTAHI KINGI, CHRIS ZHANG, BRUNO GASPERINI, AARON HEUSER, MINH HUYNH, JAMES MOORE

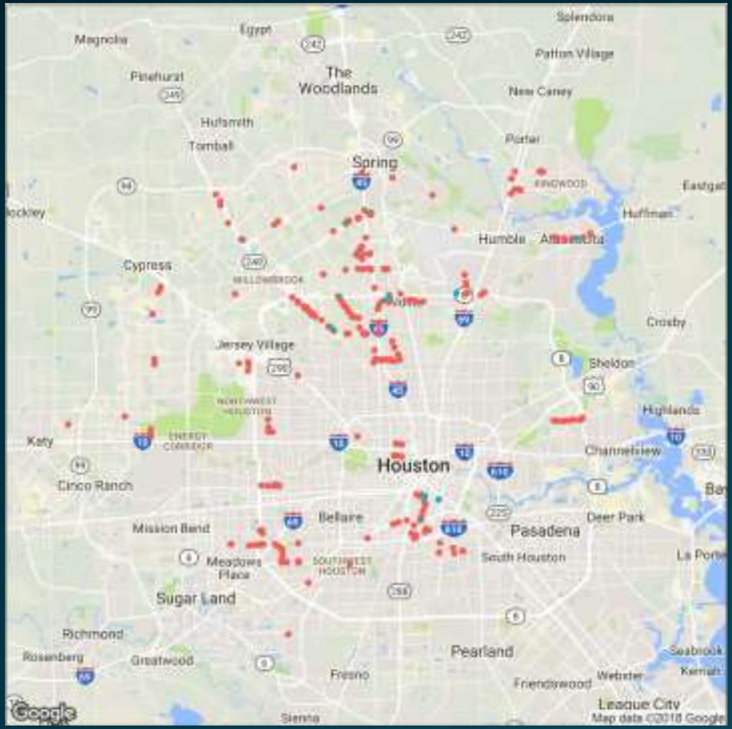
Introduction to Predictive Policing

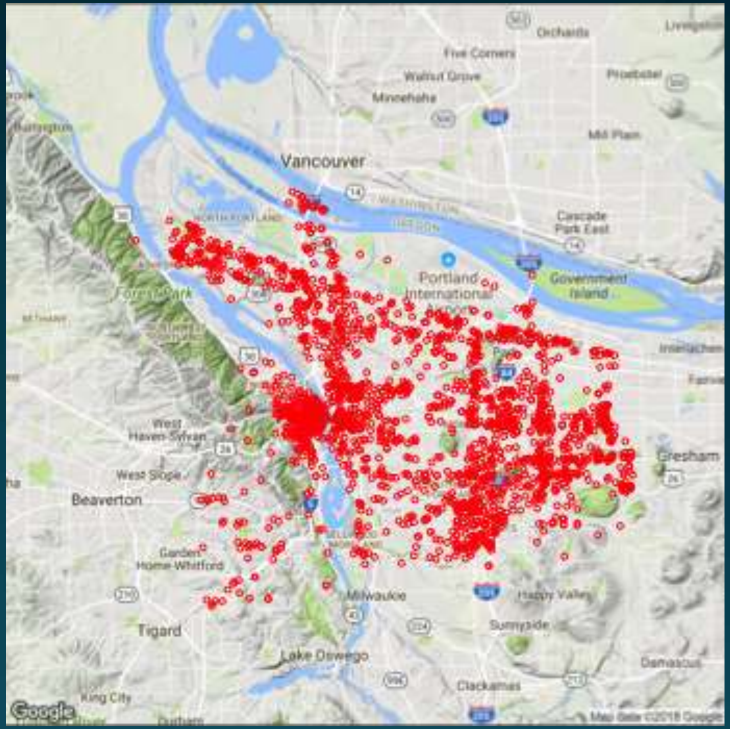
- **Crime is highly clustered - in time and space** (Sherman et al. 1989; Budd 2001; Clark and Eck 2005)
  - **⇒ Random police patrolling is ineffective**
  - **⇒ modern policing concentrates resources in high risk "hotspots"**
- **Law-enforcement demand + More datasets + Methods/comp. advances**

  = Extremely active development of predictive policing techniques
- **Predictive policing: the application of analytical techniques to identify promising geographical targets for police intervention.**

Introduction to Predictive Policing

- **Applications**
  - **Optimal Inspection Regimes**
  - **Reactive vs Preventative**
- **Clustering:**





Literature

- Many methods
  - Spatial kernel density smoothing (Johnson et al. 2009; Gorr and Lee 2015)
  - Risk terrain modelling (Caplan et al. 2010)
  - Natural language processing (Wang et al. 2012)
  - Self exciting point processes (Mohler et al. 2011; Rosser and Cheng 2016)
  - Marked Point Process (Mohler 2014)
  - Deep neural networks (Kang and Kang 2017, Duan et al. 2017)
  - Agent based crime forecasting (Malleson and Birkin (2012)

Literature

- **Many methods**
  - Spatial kernel density smoothing (Johnson et al. 2009; Gorr and Lee 2015)
  - Risk terrain modelling (Caplan et al. 2010)
  - Natural language processing (Wang et al. 2012)
  - Self exciting point processes (Mohler et al. 2011; Rosser and Cheng 2016)
  - Marked Point Process (Mohler 2014)
  - Deep neural networks (Kang and Kang 2017, Duan et al. 2017)
  - Agent based crime forecasting (Malleson and Birkin (2012)
- **Many implementations**

PREDPOL The Predictive Policing Company

HunchLab

CIVICSCAPE

Literature

- **Many methods**
  - Spatial kernel density smoothing (Johnson et al. 2009; Gorr and Lee 2015)
  - Risk terrain modelling (Caplan et al. 2010)
  - Natural language processing (Wang et al. 2012)
  - Self exciting point processes (Mohler et al. 2011; Rosser and Cheng 2016)
  - Marked Point Process (Mohler 2014)
  - Deep neural networks (Kang and Kang 2017, Duan et al. 2017)
  - Agent based crime forecasting (Malleson and Birkin (2012)
- **Few evaluations**

  "there is little consensus in academic circles on how best to assess and compare a new method and systematic evaluation is virtually absent in operational environments" – Adepeju et al. (2016)

Today

- Describe 4 (event-based) crime forecasting techniques
  - State of the art spatio temporal marked point process method (Mohler 2014)
  - 3 simplified versions Simple Crime Counts, Hawkes Process, Spatial Model
- Train models on crime data from Portland, Oregon for April-May 2017
- Predict daily crime (calls) to inform daily operations.
- Evaluate comparative performance across multiple days and crimes

Today

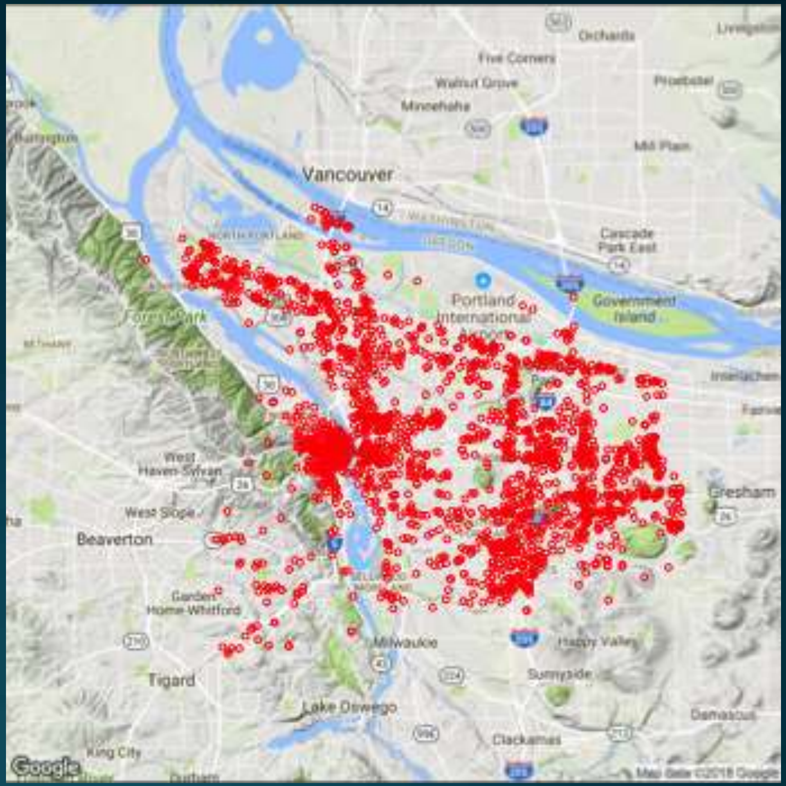
- Describe 4 (event-based) crime forecasting techniques
  - State of the art spatio temporal self exciting point process method (Mohler 2014)
  - 3 simplified versions Simple Crime Counts, Hawkes Process, Spatial Model
- Train models on crime data from Portland, Oregon for April-May 2017
- Predict crime
- Evaluate comparative performance across multiple days and crimes
- Not evaluating:
  - Techniques identifying individuals at risk of offending
  - Methods predicting perpetrators' identities
  - Algorithms predicting victims of crimes
  - Performance against other important criteria like racial bias

Data

- **Public data of reported crime occurrences in Portland, OR for April-May 2017**
- **Provided by the National Institute of Justice for Crime Forecasting Competition**
- **Example Dataset:**

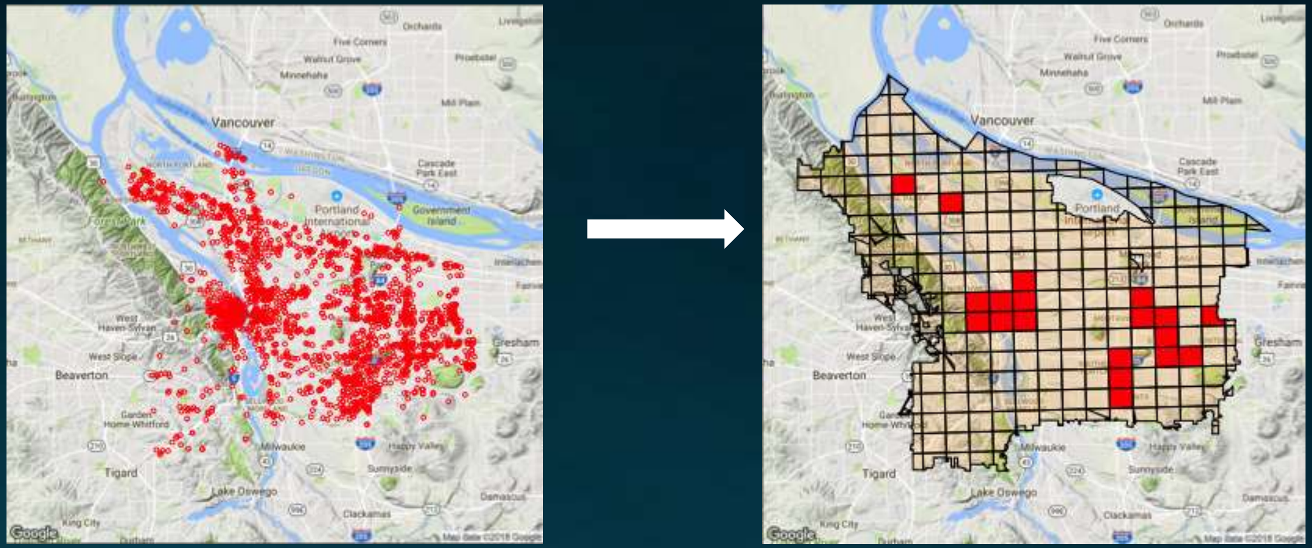
Category	Date	Latitude	Longitude
Burglary	4/5/2017	45.538723	-122.477039

Burglaries April 2017



1. Simple Counts

- **Split geography into grids, $g \in G$**
- G chosen as 600ft x 600ft grids
    - Literature
    - Realistic policing requirements
- $C_g = \sum_i crime_{i,g}$

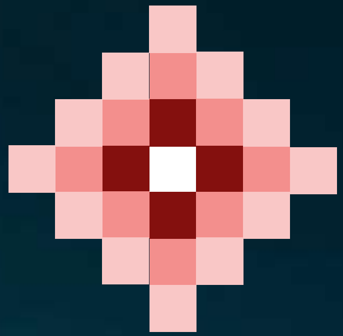


- **Pros: Simple**
    - **What (most) PD do; used as benchmark here**
- **Cons: Does not account for spatial or temporal dimension and different crime types**

2. Spatial Model (Arraiz et al. 2010)

- **Seemingly Unrelated Regression among 4 categories with Spatial Weight**

$$y_t^x = \alpha^x + \tau y_{t-1}^x + \rho W y_t^x + \sum_{z \in \{b,m,s,o\},\ z \neq x} \beta^z y_t^z + \varepsilon_t^x, \qquad x \in \{b, m, s, o\}$$



- **Weight matrix:** k nearest neighbors (k = 24)
- **IV:** y_{t-1}^c for y_t^c , c $\in \{b, m, s, o\}$
- **Split data in 2 equal-sized windows (15-day) to estimate parameters**
- **Flag hotspots based on** $(\hat{y}_t^b, \hat{y}_t^m, \hat{y}_t^s, \hat{y}_t^o)$
- **Pros:** spatial and (some) temporal dimension and different crime types
- **Cons:** temporal dimension in a restrictive way (one-period lag) and linear specification

3. Hawkes (1971) Process

- **Extension of simple Poisson λ process.**

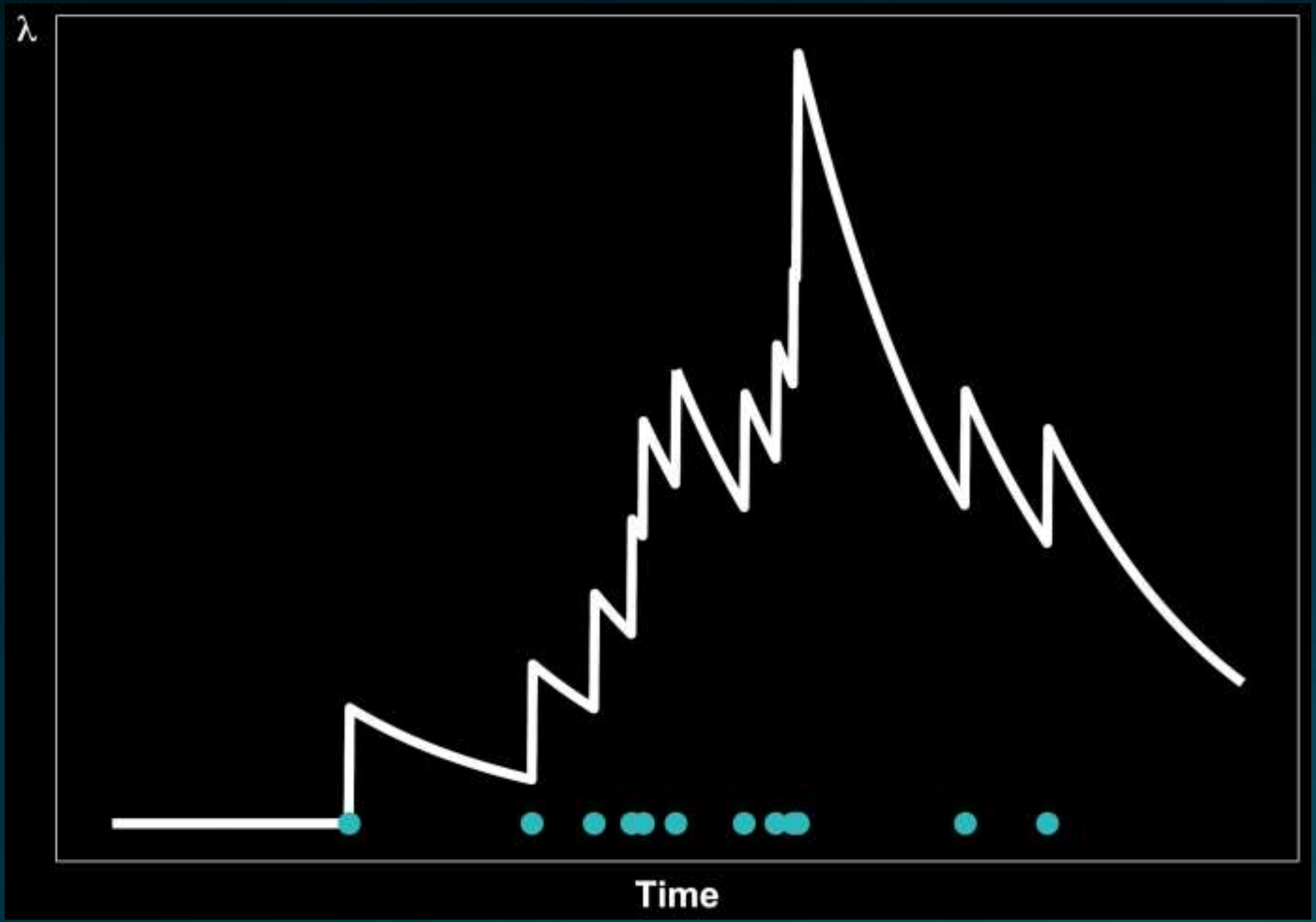
$$\lambda_s(t) = \mu_s + g_s(t)$$

- μ_s: **Background rate → structural difference across grids**
- $g(\cdot) = \sum_{t_i<t} \alpha_s e^{-\beta_s(t-t_i)}$: **Triggering function → near-repeat time effects**

- **Pros: reflects criminology crime clustering explanations like "broken window" theory**
- **Cons: ignores spatial dimension**

3. Hawkes (1971) Process

- **Self-exciting process ⇒ Clustering**



4. Mohler (2014)

- **Marked Point Process**
  - Spatial and temporal dimension
  - Developed for earthquake modeling (Daley and Vere Jones 1988)
  - Also, different crime types $M = 1, 2, \ldots, N_c$

- **Crime intensity modeled as:**

$$\lambda(x, y, t) = \mu(x, y) + \sum_{t>t_i} g(x - x_i, y - y_i, t - t_i, M_i)$$

  - $\mu(\cdot)$: Background rate → stationary component (intrinsic differences across "grids")
  - $g(\cdot)$: Triggering function → near repeat effects (space, time, and crime types)

4. Mohler (2014)

- **Triggering function:**

$$g(x,y,t,M) = \theta(M)\omega exp(-\omega t) \times \frac{1}{2\pi\sigma^2} exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

  - Exponential decay in time: ω determines the timescale
  - Gaussian in space: σ controls the length scale

- **Background rate:**

$$\mu(x,y) = \sum_{t>t_i} \frac{\alpha(M)}{T} \frac{1}{2\pi\eta^2} \times exp\left(-((x-x_i)^2+(y-y_i)^2)/2\eta^2\right)$$

4. Mohler (2014)

- **11 Parameters to estimate:** $(\omega, \sigma, \eta, \theta^b, \dots, \theta^o, \alpha^b, \dots, \alpha^o)$

- **Expectation-Maximization (EM) algorithm:**
  - Each crime generated by one of the mixture kernels (with certain probabilities)
  - Convergence: probabilities are proportional to the value of the kernel at the crime space time location relative to the sum of all kernels at the crime location
  - E step: determine the probabilities that event i trigger crime j
  - M step: given probabilities from E step, updates parameters
  - For a given initial guess, EM algorithm updates the probabilities and the parameters until convergence

4. Mohler (2014)

- **Pros:**
  - Spatial and temporal dimension and different crime types
  - Models clustering

- **Cons:**
  - Complex → functional form assumptions and computational costs

Model Hierarchy

Spatial

SEM:

$\mu(s) + g(s - s_i)$

Mohler:

$\mu(s) + g(t - t_i, s - s_i)$

Simple Counts:

$\mu(s)$

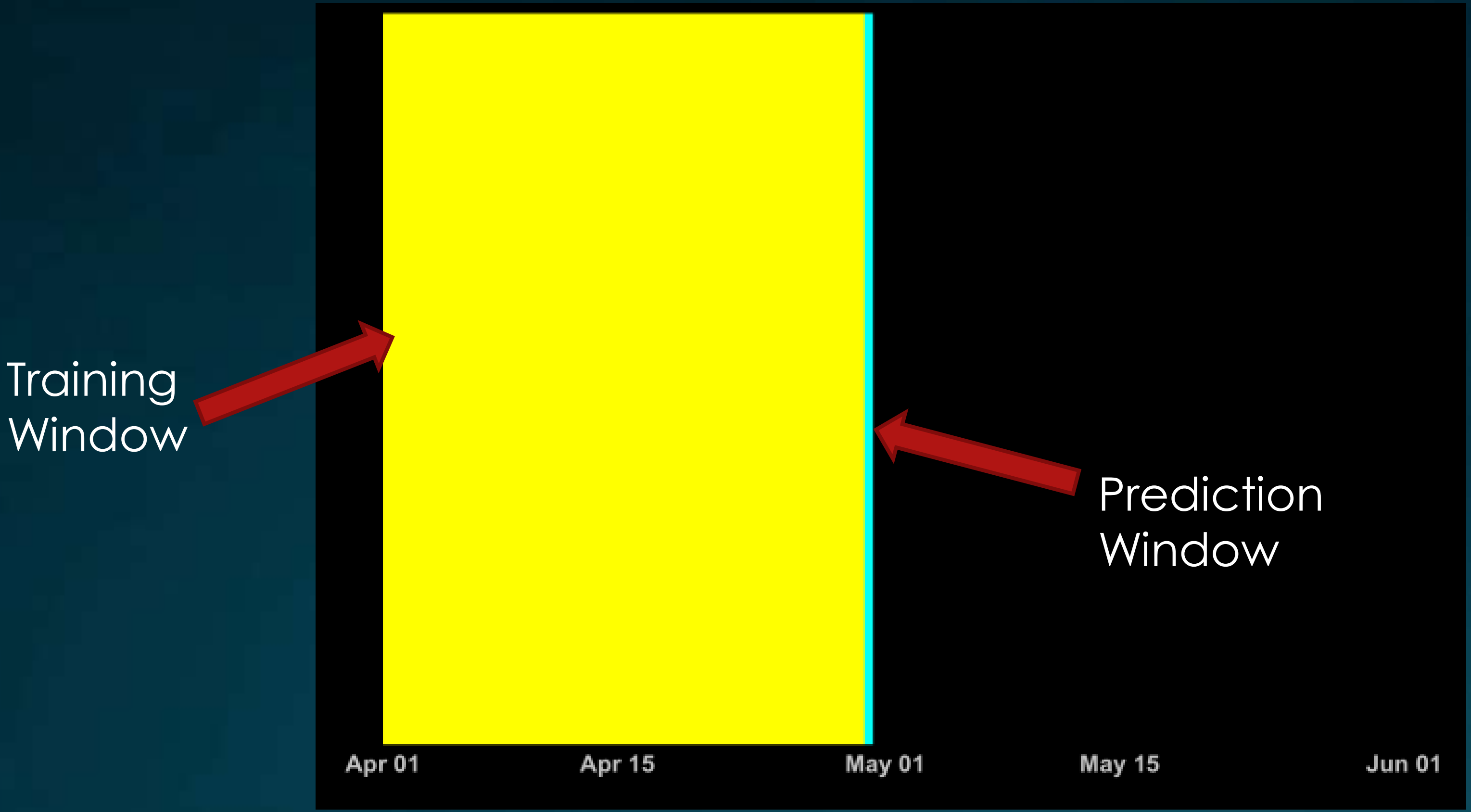
Hawkes Process:

$\mu(s) + g(t - t_i)$

Temporal

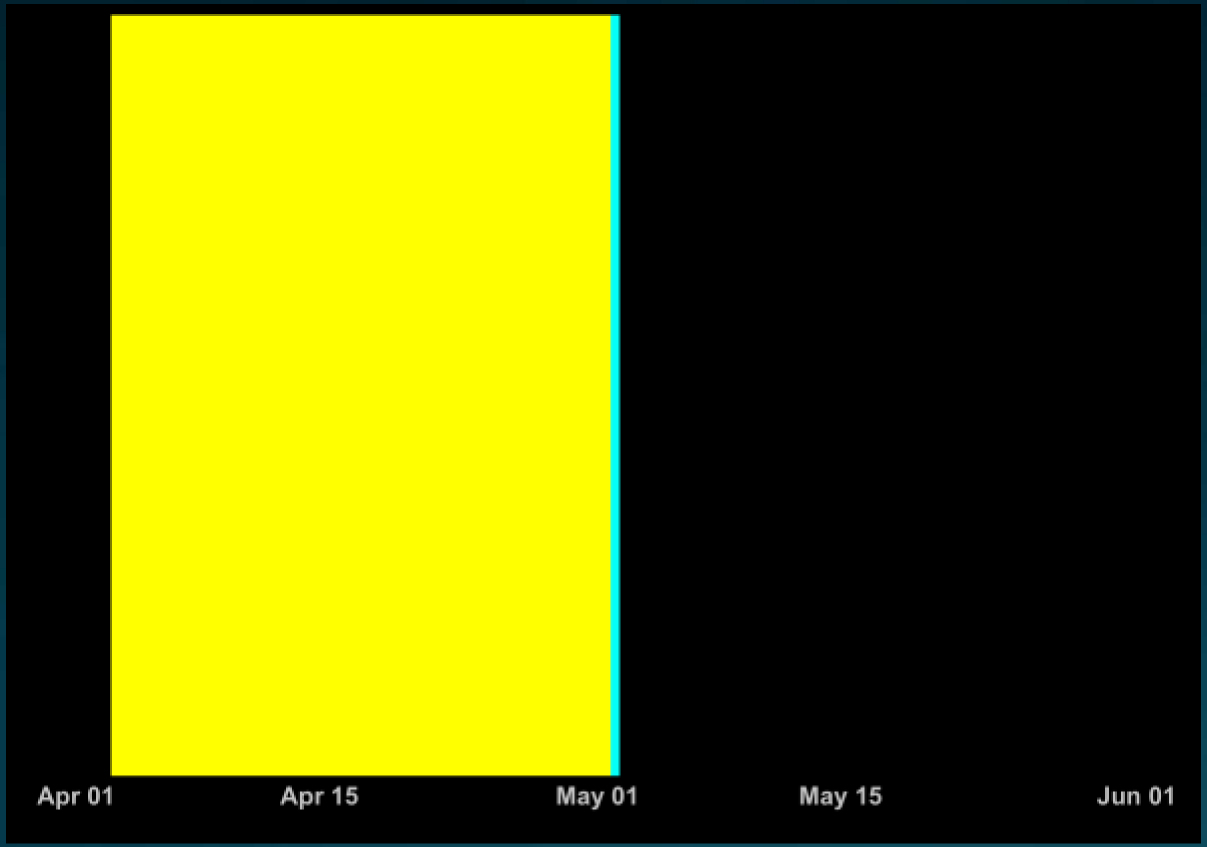
Forecast Approach

- **1 day prediction window (consistent with police practice) through May 2017**
- **30-day rolling training window**



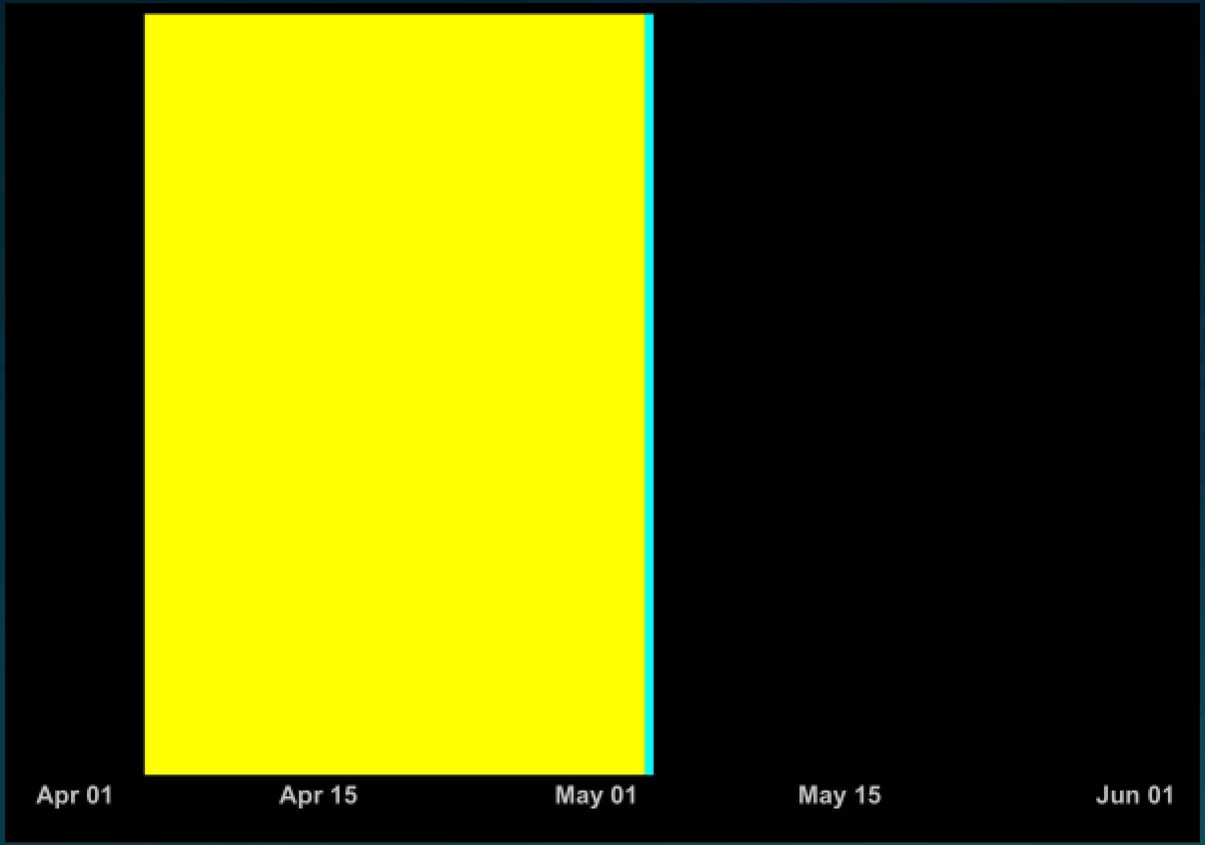
Forecast Approach

- **1 day prediction window (consistent with police practice) through May 2017**
- **30-day rolling training window**



Forecast Approach

- 1 day prediction window (consistent with police practice) through May 2017
- 30-day rolling training window



Forecast Approach

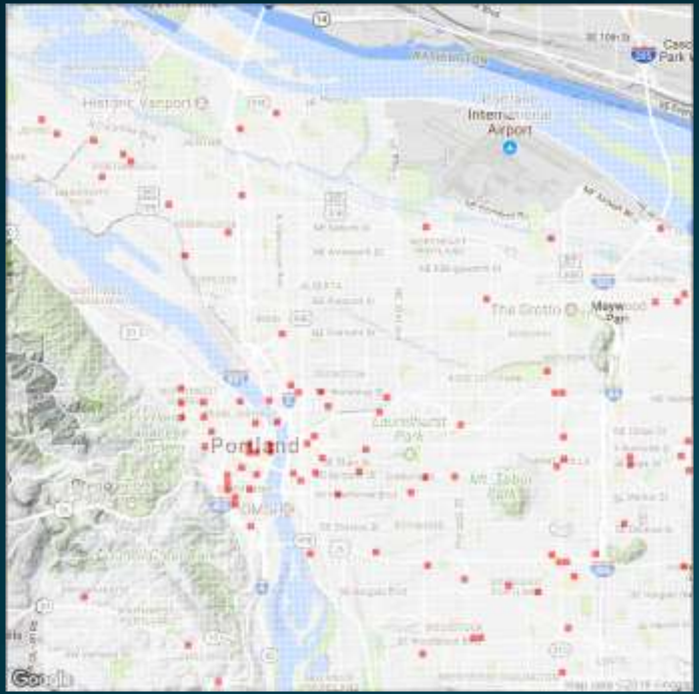
- 1 day prediction window (consistent with police practice) through May 2017
- 30-day rolling training window

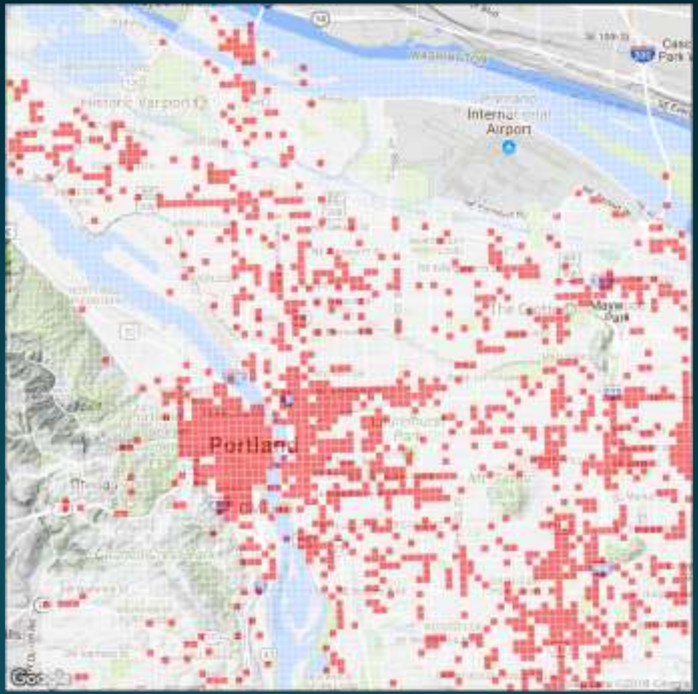
Forecast Approach

- **1 day prediction window through May 2017**
- **30-day rolling training window**
- **600ft x 600ft grids ($\approx 12,000$ grids)**

Forecast Approach

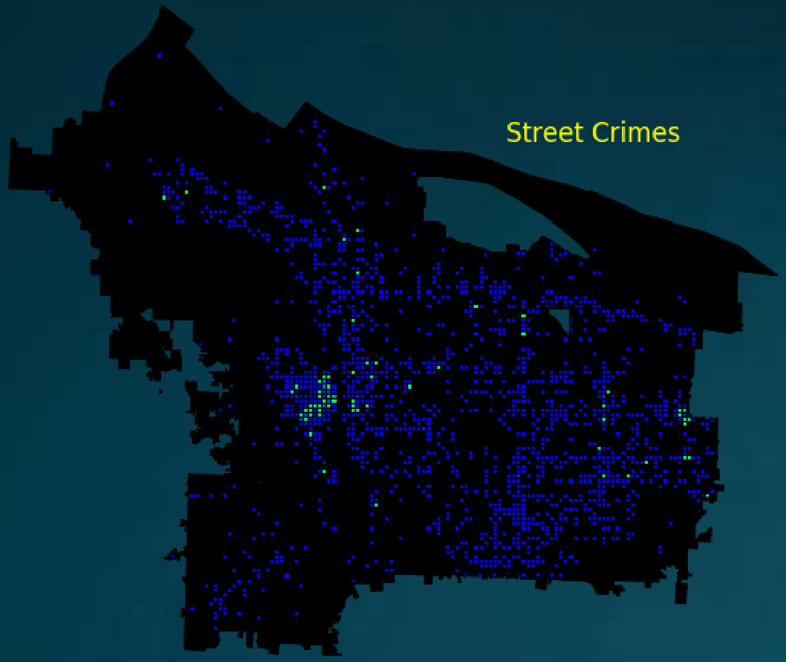
- 1 day prediction window through May 2017
- 30-day rolling training window
- 600ft x 600ft grids ($\approx 12,000$ grids)
- 2 coverage areas: 1% ($\approx$ 120 hotspots) and 15% ($\approx$1,800 hotspots)

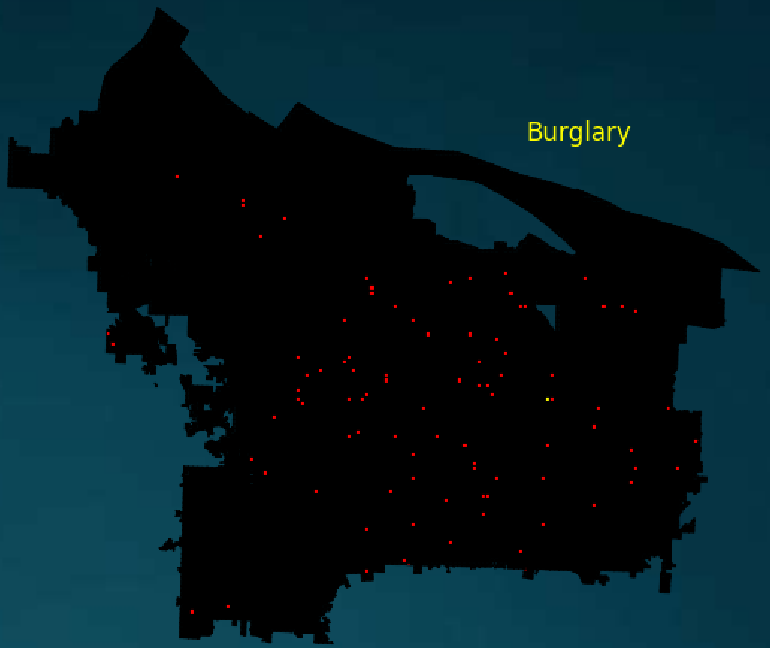




Forecast Approach

- 1 day prediction window through May 2017
- 30-day rolling training window
- 600ft x 600ft grids ($\approx 11,000$ grids)
- Two crime types: Street Crime, Burglary





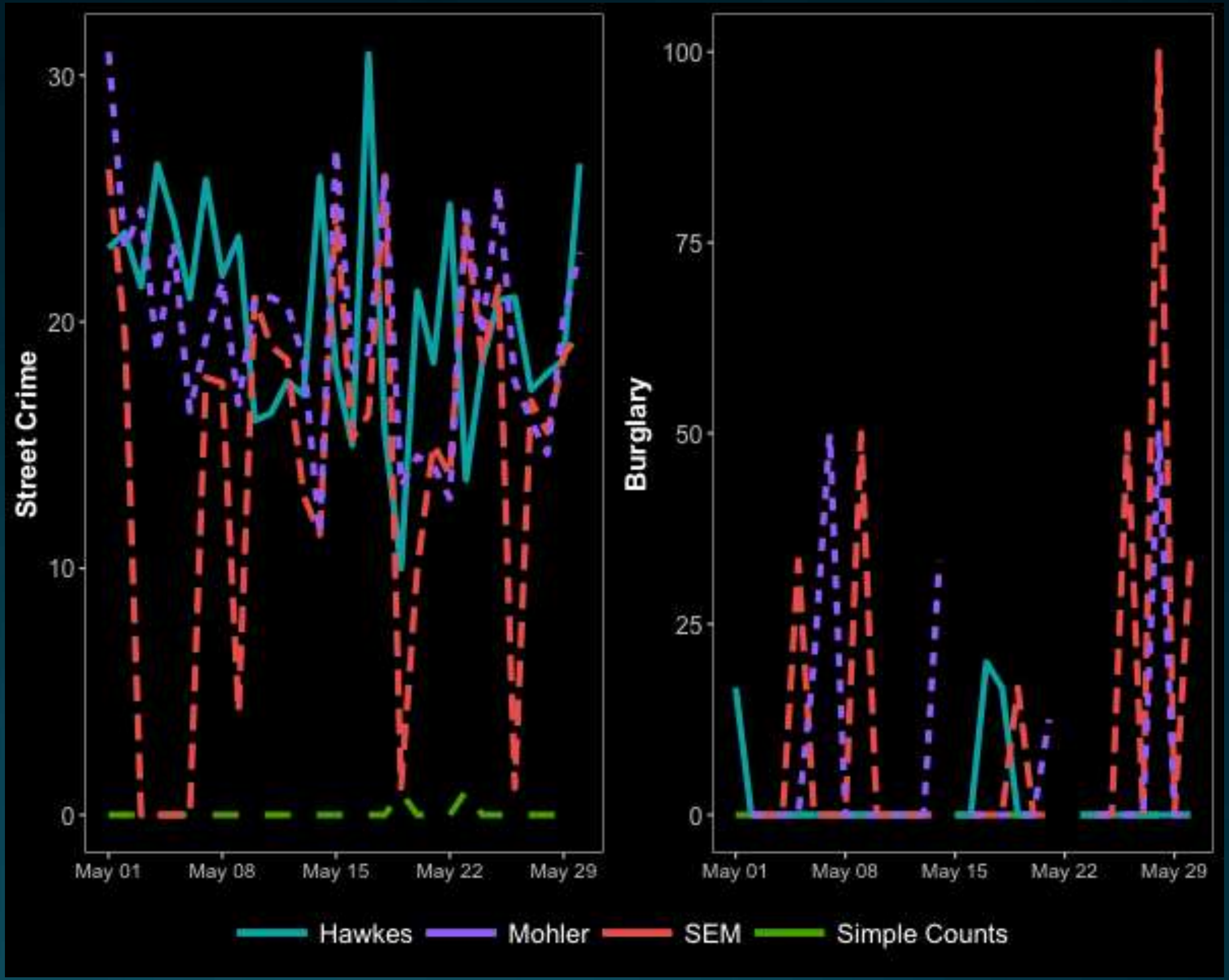
Forecast Metrics

- **Traditional Forecast/Prediction Evaluation Methods: MAE, RMSE**
  - Evaluate entire space and time domain ⇒ unsuitable for sparse (many zero) crime data
- **Hit Rate:** $HR = \frac{n}{N}$
  - Simple, intuitive. Can be artificially inflated by increasing hotspot size, limited practical use for policing.
- **Predictive Accuracy Index:** $PAI = \left(\frac{n}{N}\right) / \left(\frac{a}{A}\right)$
  - Crime density in hotspots / crime density over the whole region. Hit Rate that accounts for coverage area
- **Prediction Efficiency Index:** $PEI = \frac{n}{n^*}$
  - Performance of forecast compared to optimal (ex post) solution
  - n^*: maximum number of crimes that can be captured within k grids, where k is the number of hotspots.

Graphical Results

PAI - 1% Coverage

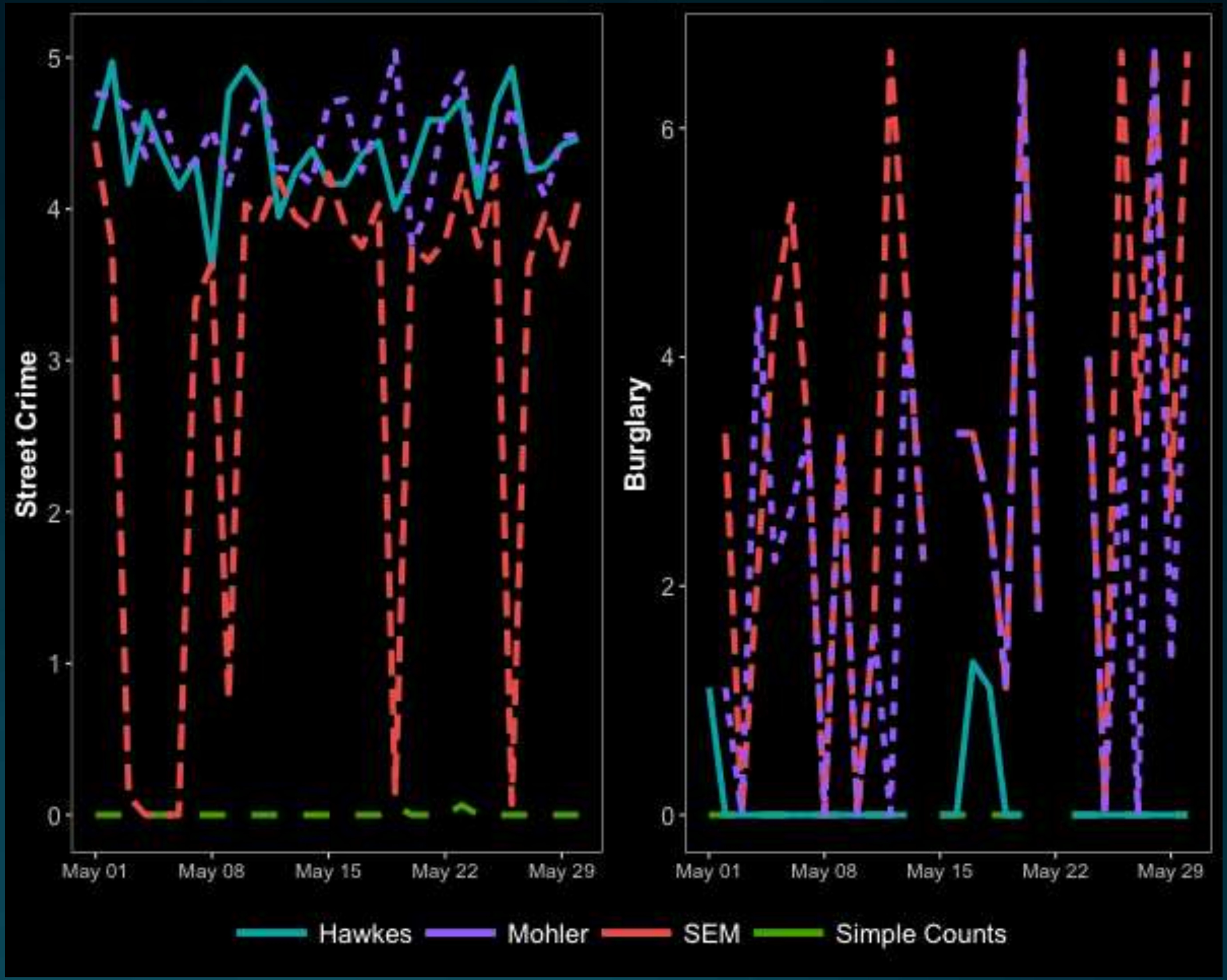
- $PAI = \frac{n}{N}\frac{1}{0.01}$



Graphical Results

PAI - 15% Coverage

- $PAI = \frac{n}{N}\frac{1}{0.15}$
- Smaller magnitude than 1% coverage
- Conjecture:
  - “Emotional” crimes better predicted by temporal models
  - “Planned” crimes better predicted by spatial models.



Analytical Results

1% Coverage

		PAI	
Crime Type		Mean	SD
Street Crime	Simple Counts	0.063	0.239
	SEM	14.178	8.377
	Hawkes	**20.370**	**4.552**
	Mohler	19.749	4.653
Burglary	Simple Counts	0.000	0.000
	SEM	**10.900**	**24.006**
	Hawkes	1.975	5.717
	Mohler	6.377	14.977

15% Coverage

		PAI	
Crime Type		Mean	SD
Street Crime	Simple Counts	0.004	0.016
	SEM	3.030	1.631
	Hawkes	4.407	0.317
	Mohler	**4.456**	**0.293**
Burglary	Simple Counts	0.000	0.000
	SEM	**3.299**	**2.197**
	Hawkes	0.132	0.381
	Mohler	2.461	1.960

Sign Test

- Comparing mean values hides inherent variability in results, especially with techniques designed to improve upon averaging methods.
- **Wilcoxon signed-rank test**
  - **Consider values as time series**
  - **Assumption: Difference in predictive accuracy between methods is independent of underlying crime rate ⇒ time series of differences are i.i.d.**

$$W = \sum_{t=1}^{T} sgn(y_{2,t} - y_{1,t}) \cdot R_t$$

- $T = 30$, $y_{i,t}$ = relevant accuracy measure (HR, PAI, or PEI) of method i for day t, R_t = rank of the difference and $sgn(\cdot)$ is the sign function

Formal Results

		1% Coverage		15% Coverage	
Crime Type	Method	> Simple Count	> Mohler	> Simple Count	> Mohler
Street Crime	Simple Counts	-	NS	-	NS
	SEM	***	NS	***	NS
	Hawkes	***	NS	***	NS
	Mohler	***	-	***	-
Burglary	Simple Counts	-	NS	-	NS
	SEM	*	NS	***	*****
	Hawkes	*	NS	*	NS
	Mohler	*	-	***	-

* statistical significance at 10% level and 1% level, respectively.
NS Not statistically significant

Conclusion/Extensions

- Formalize model connections
- Expand list of models
- Extend training windows (perhaps with HPC)
- Formally train (rather than estimate) model parameters
- Extension of Mohler (2014):
  - Moving away from Gaussian assumption for triggering function
  - Distributions that allow for rare events (i.e., fat tails)