Transferable Spatial Signatures for Crime Forecasting

Jason R. Dalton

7921 Jones Branch Rd. Suite 600
McLean, VA 22102
jason.dalton@spadac.com

Presented to the Ninth Crime Mapping Research Conference;
March 28-31, 2007
Pittsburgh, PA

Problem Description

Given a geographic space, we have an event point process occurring in that region. The objective is to identify the information layers that are responsible for the event site selection, understand their relative importance, and construct a likelihood contour for a future event.

Motivation

A reliable forecast method for specific modalities of criminal activity would improve resource allocation for patrols, sensors, and crime reduction strategies. Improved resource allocation results in an enhanced ability to detect, classify, and mitigate serious crime before they endanger the population.

Defining a Spatial Decision Process

To commit an attack, one must consider the opportunity space in which to act.

Some subset of the opportunity space is actually selected by an attacker.

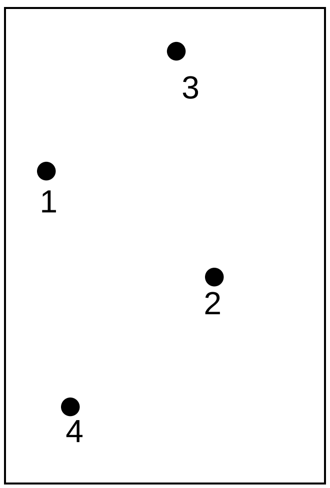
These spaces can be well formulated probabilistically, and appropriate models can be fit using empirical observation.

When observed at the micro-geographic scale (city block size), the variation in the environment is high – nonhomogeneus background.

We examine the relative distribution of the opportunity space to the choice space to discover an attackers preferences.

Spatial Feature Extraction

Mailboxes

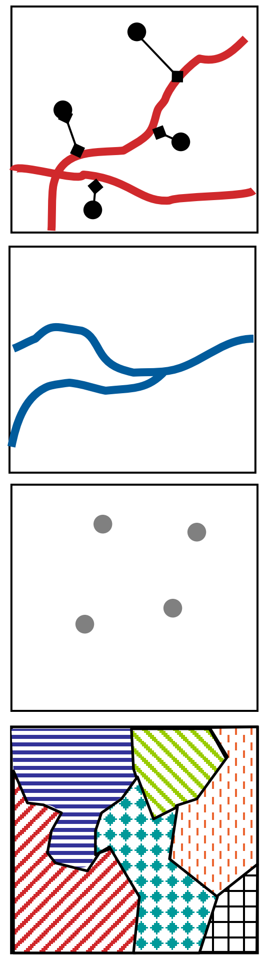


Roads

Rivers

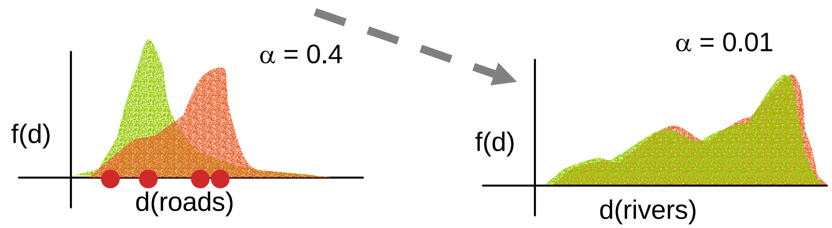
Hydrants

LandUse



	R	R	7	E
1	1.2	2.2	.5	A
2	1.3	4.3	.5	C
3	1.7	3.2	.5	H
4	6	2.5	.5	P

$$P[X = x_i \mid] = \sum_{j=1}^{p} \alpha_j \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-d/\sigma^2}$$



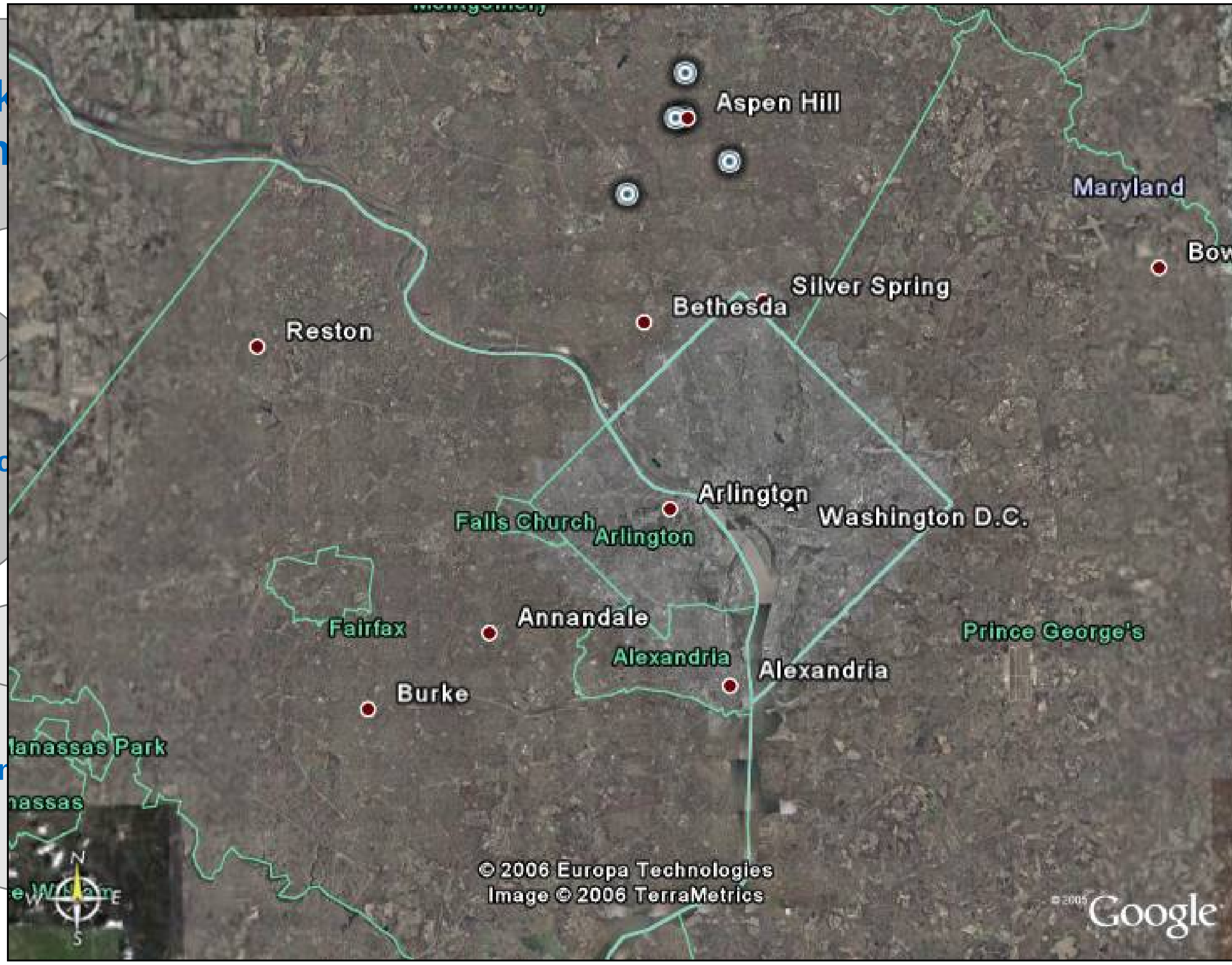
$$[P(object)] = \sum_{j=1}^{p} \alpha_j \log\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}} e^{-d_i/\sigma^2}\right] - \log\left[\frac{1}{m}\sum_{i=1}^{m}\frac{1}{\sqrt{2\pi\sigma^2}} e^{-d_i/\sigma^2}\right]$$

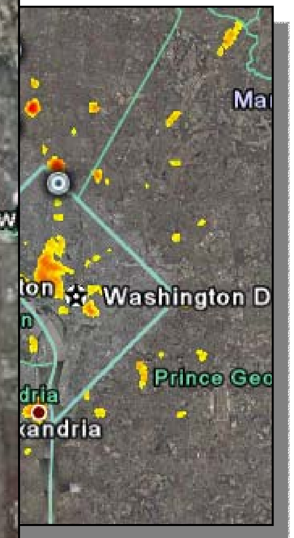
Beltway Sniper Example

5 Attack
Location

Socio-Economic

Der



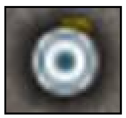


igence for

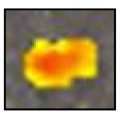
vernment.

SPADAC

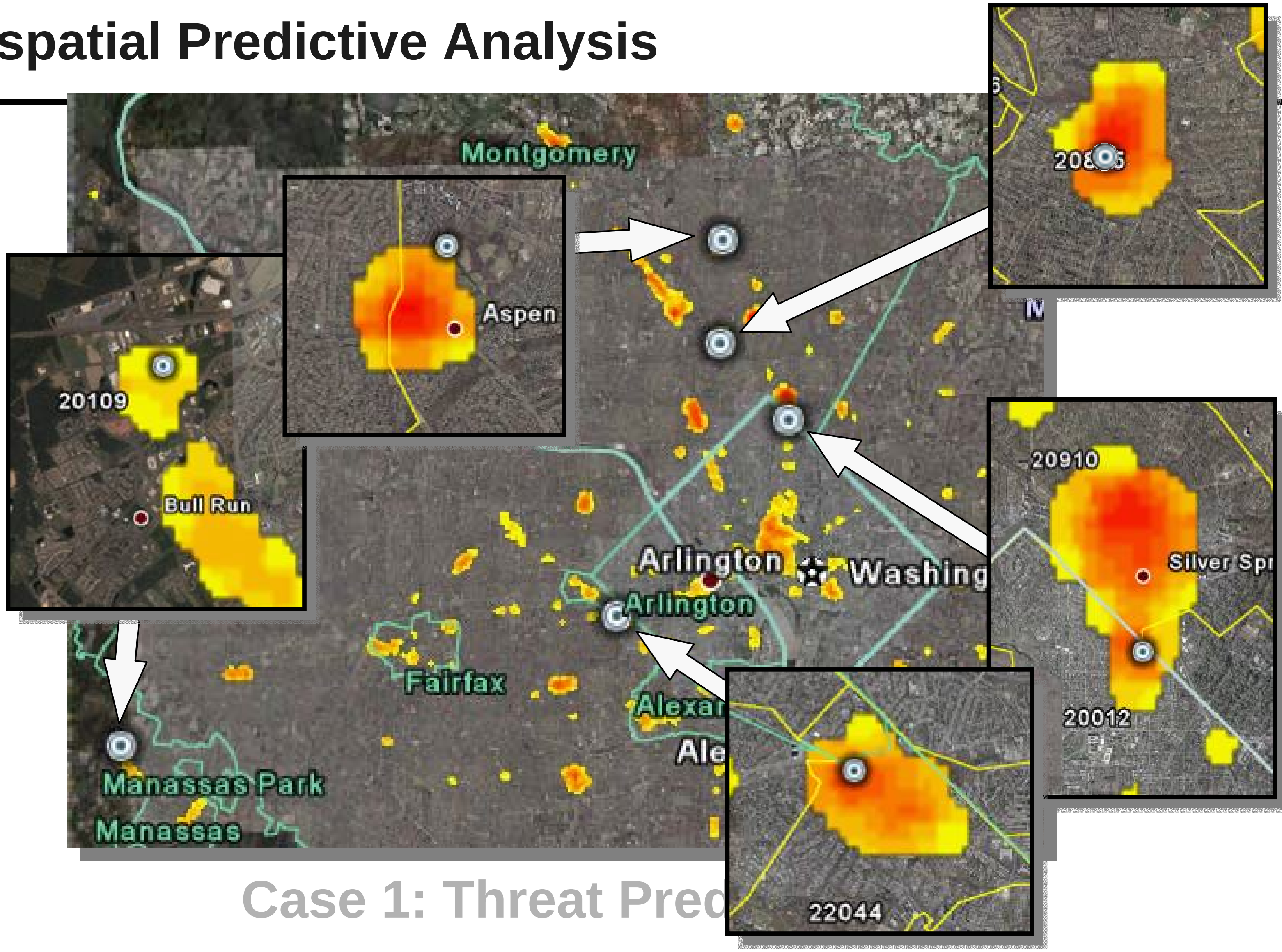
Geospatial Predictive Analysis



Attack Location



Highest Likelihood Areas



Case 1: Threat Pred

The Data

The environment data: g_{ij} is the j^{th} factor at l_i

$$\mathbf{G} = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ g_{p1} & g_{p2} & \cdots & g_{pm} \end{pmatrix}$$

The event data

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$$

A probabilistic model

- $y = (y_1, \ldots , y_m)^T$: a vector of m-factors at a particular location
- ν : the density function of the environment of y
- f : the density function of the events of y
- Define the selection bias function:

$$\delta(y) = \frac{f(y)}{\nu(y)}$$

Probabilistic justification

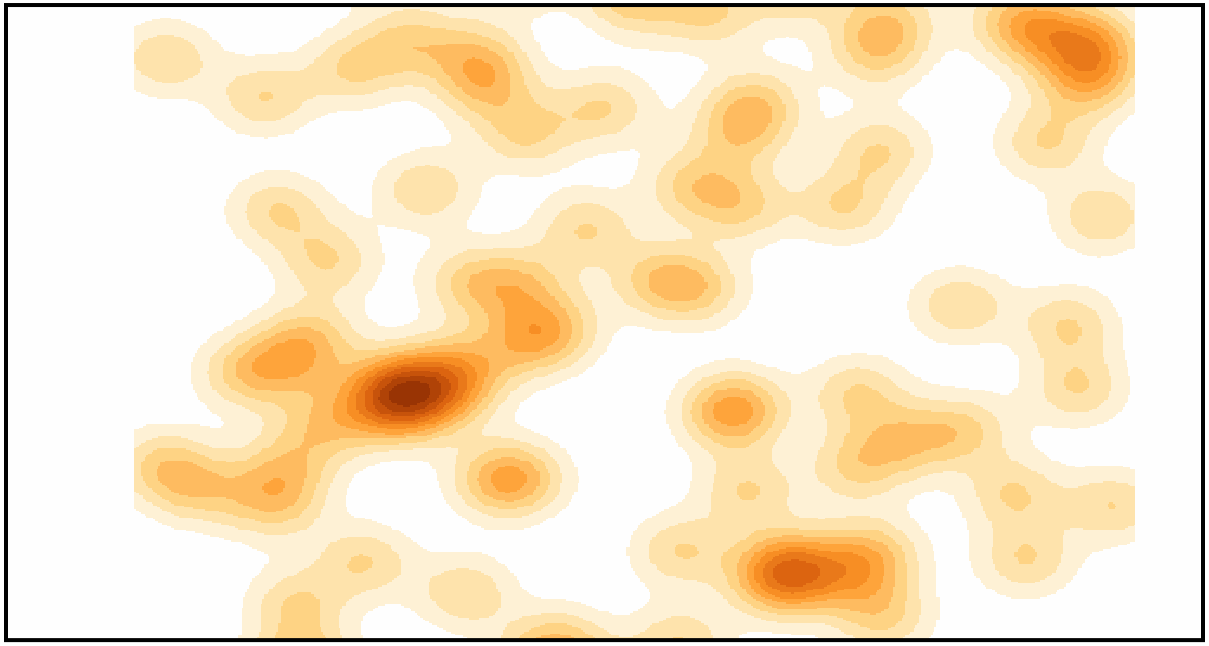
- $P(\text{ event} \mid y\ )$: conditional probability of an event at a location with factor value y.
- Given two locations with factors y and u the event location is chosen with the odds $P(\text{ event} \mid y\ ) : P(\text{ event} \mid u\ )$
- Probability of choosing a location with factor value y at random is $\nu(y)$.
- Given an event, the conditional probability of the event occurring at a location with factor value y is

$$\frac{P(\text{ event} \mid y\ )\nu(y)}{\int P(\text{ event} \mid y\ )\nu(y)dy} := f(y)$$

- δ is the normalized version of $P(\text{ event} \mid y\ )$.

Likelihood Surface

- Likelihood of a future event at different locations are proportional to $\delta(y)$ for the factor values y at those locations.
- Evaluated along a regular grid, this generates a likelihood contour of future events on that geographic region.



Variable selection

We evaluated several methods for estimating the importance of a geographic variable, defined as the magnitude of the selection bias function.

Recall:

$$\delta(y) = \frac{f(y)}{\nu(y)}$$

Entropy measure: $\int \ln(\delta)\,\delta\,d\nu$

Found through simulation to be the most robust estimator when δ is near 1

Hellinger Distance: $\frac{1}{2}\int(\sqrt{\delta}-1)^2 d\nu$

L_2 Distance: $\int(\delta-1)^2 d\nu$

L_1 Distance: $\frac{1}{2}\int|\delta-1|d\nu$

SPADAC

Skewness of selection

How different is the event selection process from an unbiased selection ?

- Entropy measure

$$a = \int \ln(\delta(y))\delta(y)\nu(y)dy$$

- a >= 0 always.
- a = 0 indicates a random selection. What are the implications of this?
- The larger the value of a, the more skewed the event selection.

Plug-in estimation of δ

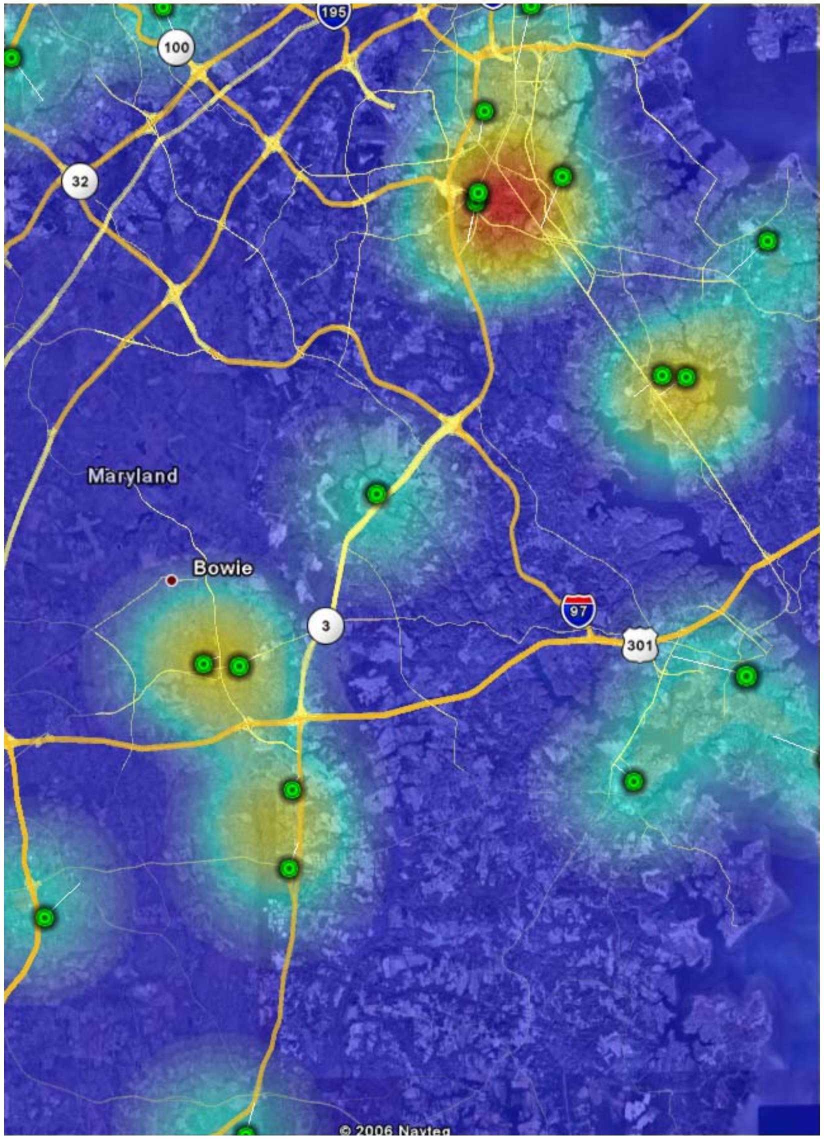
- $\hat{f}$: kernel estimate of f using the event data.
- $\hat{\nu}$: kernel estimate of ν using the environment data.
- $\hat{\delta} = \hat{f} / \hat{\nu}$
- Simple enough.
- For non-negative factors, one needs edge correction at 0.

Estimation of δ

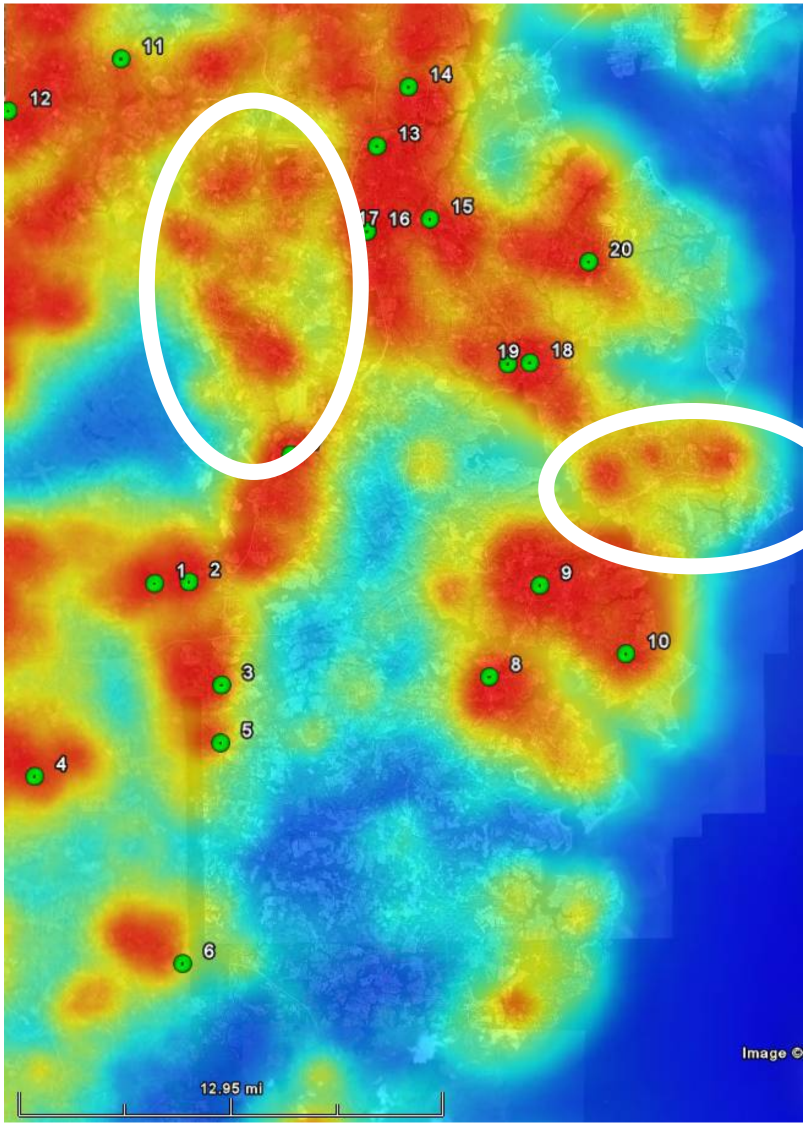
- Use only the important factors.
- Few relevant factors:
  use multivariate kernel estimate for $\hat{\delta}$.
- Many relevant factors:
  - Estimate univariate component function $\hat{\delta}_j$'s
  - Combine the component function through an *additive model*.
- Use simulated annealing method to trim the peaks and valleys of $\hat{\delta}$ or $\hat{\delta}_j$.

Differences between Spatial Assessments and Density maps

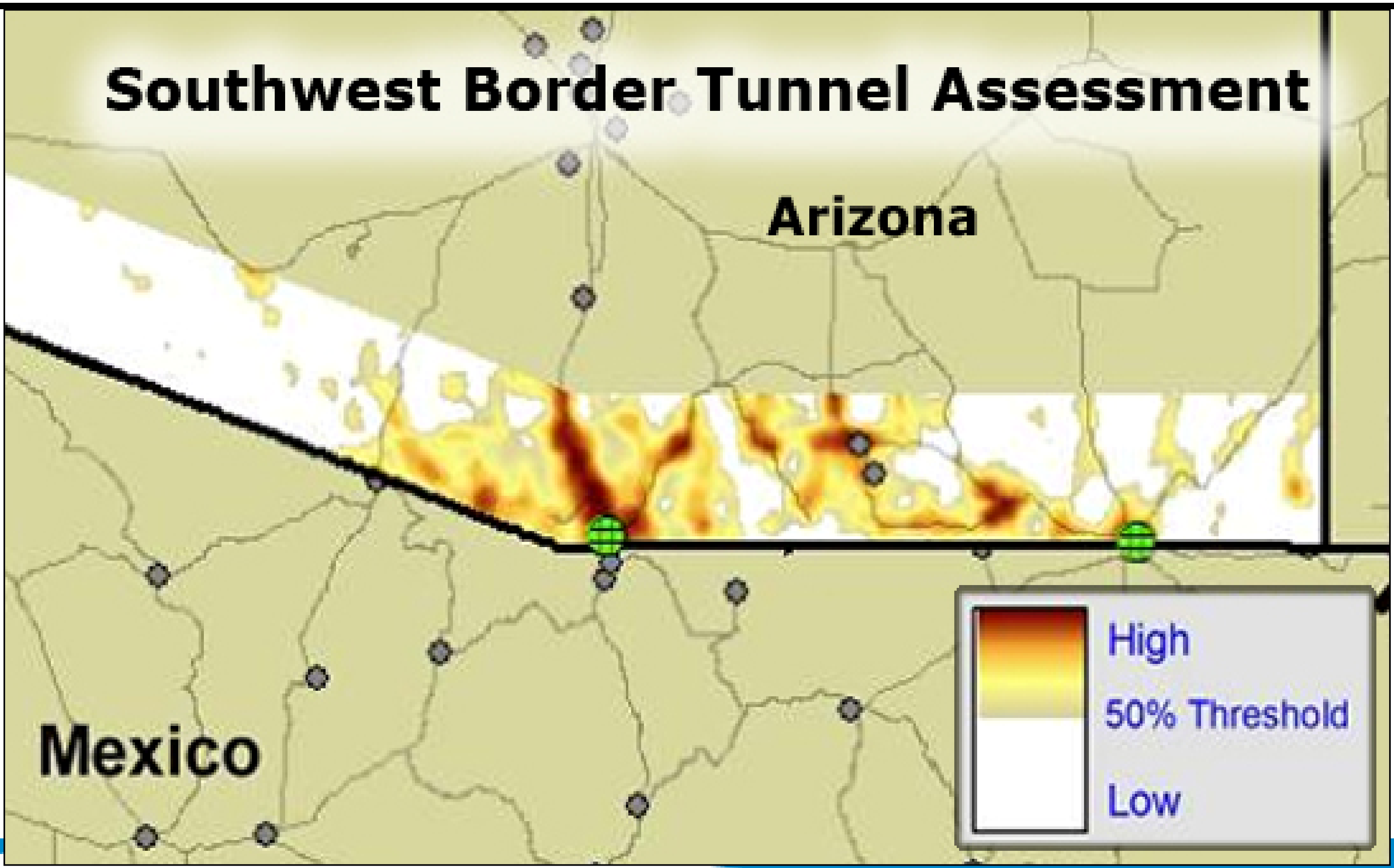
Density shows only the areas which are already known



Orion assessments show areas known and unknown, which match the features of the training points.

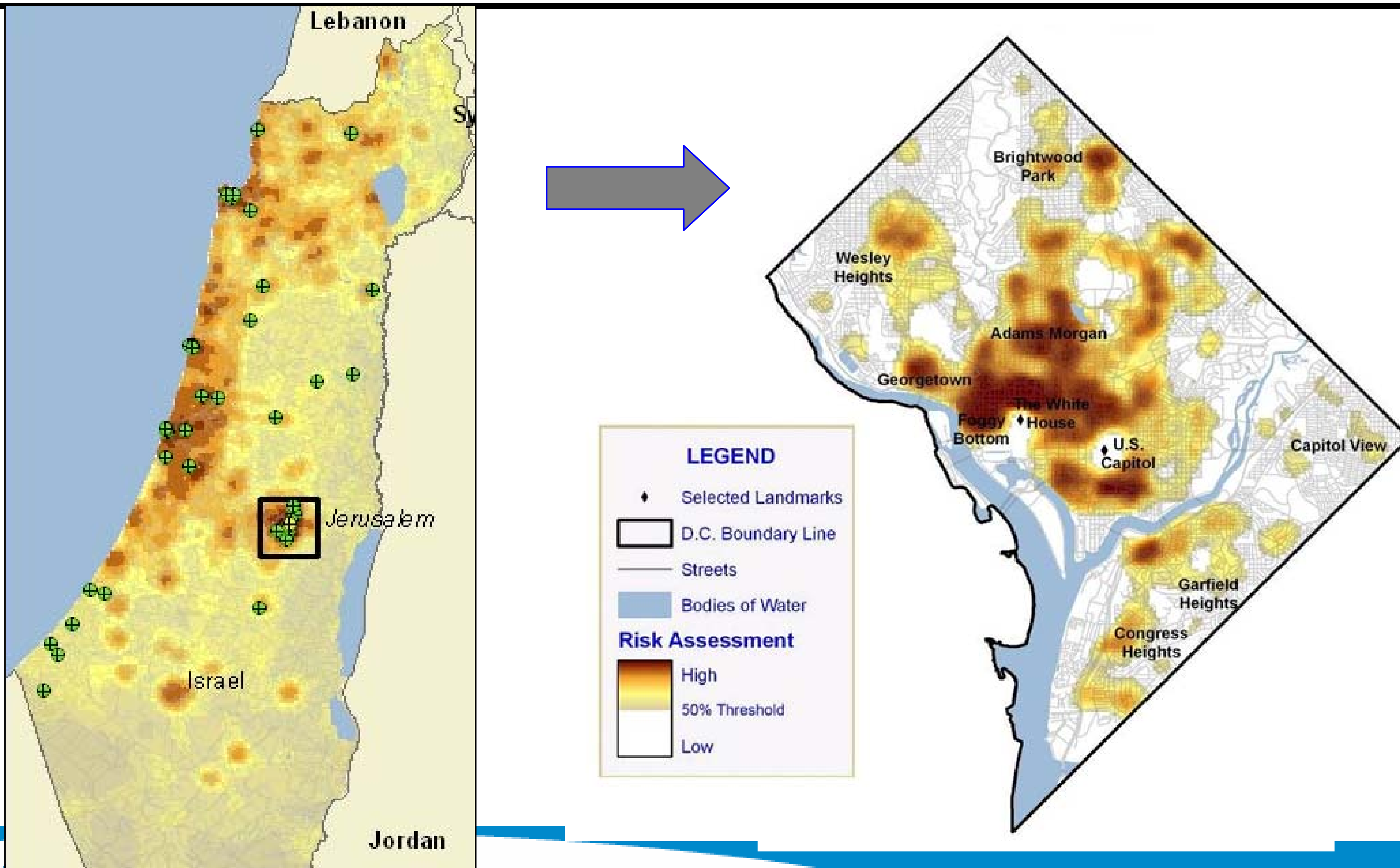


Law Enforcement Support on the US Border



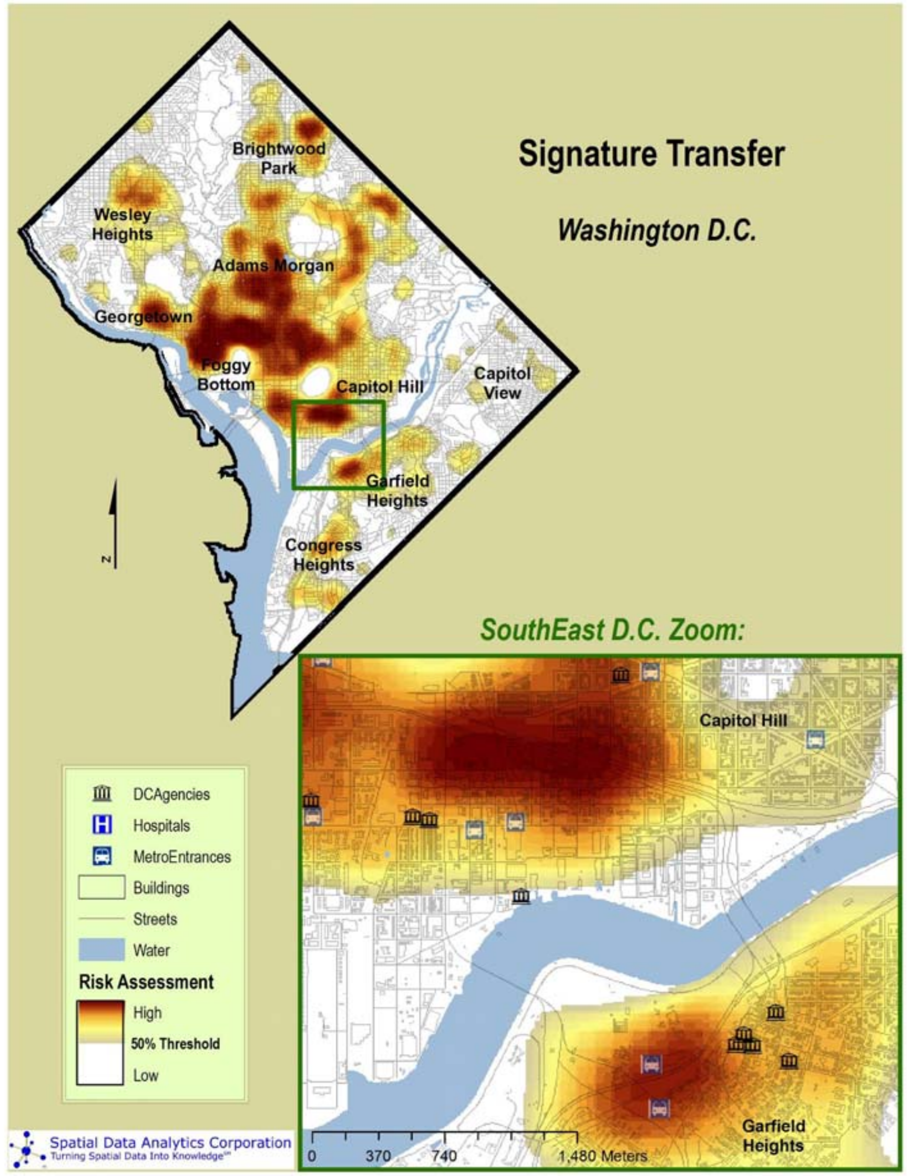
SPADAC

Projected Risk Assessment & Resource Allocation for a D.C. Terrorist Incident



Detailed look at signature transfer

- Capitol and White House are "cold-spots"
- The model observes that there are protective measures in place that make bombings infeasible.
- Each attack type has its own signature
- Presence of barriers, steps, guards, make this a "hard target"



SPADAC

Transferable Spatial Signatures for Crime Forecasting

Jason R. Dalton

7921 Jones Branch Rd. Suite 600
McLean, VA 22102
jason.dalton@spadac.com

Presented to the Ninth Crime Mapping Research Conference;
March 28-31, 2007
Pittsburgh, PA