Evaluating Crime Prediction Maps

MONSURU ADEPEJU & ROSSER GABRIEL

Big Data and Intelligent Policing

CPC Project Closing Workshop

7th June 2016

Motivation

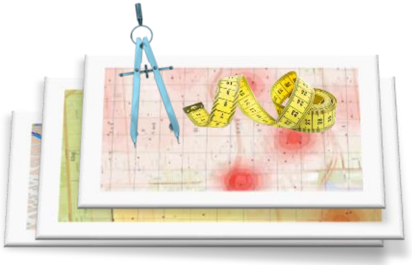
Given the increasing number of predictive hotspot methods, crime analysts often find it difficult to determine **<u>which method is most appropriate</u>** for different data scenarios.

For example,

- Which method is the best for predicting geographically constraint crime such as shoplifting?
- Which method is most robust to highly spatio-temporal sparse crime types?

Proposed solution:

A <u>systematic evaluation protocol</u> by which the performance of multiple predictive methods can be <u>assessed and compared</u>



Evaluation Framework

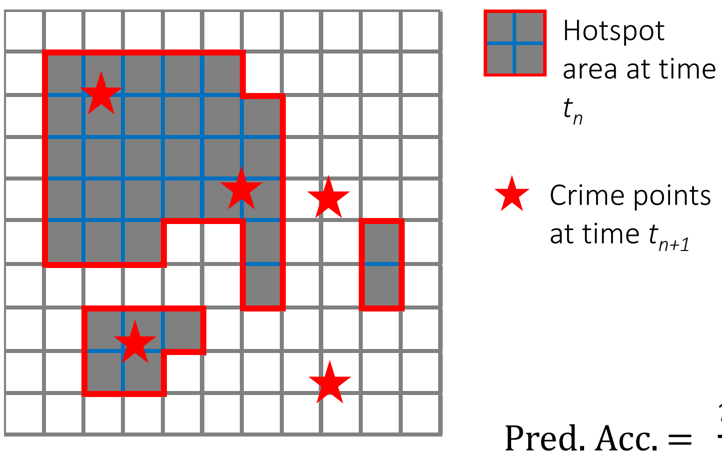
We combined measures of four different aspects of hotspot performance namely;

- Accuracy,
- Compactness,
- Variability,
- Complementarity

Evaluation metrics

1. **Predictive accuracy (hit rate)**

measures the number of crimes captured within the hotspot area



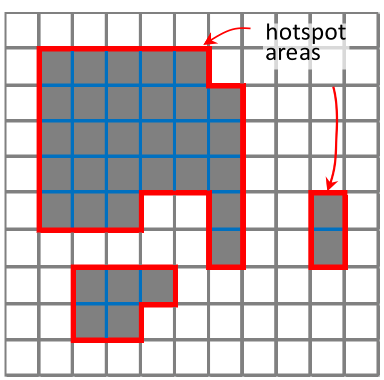
> ✓ A method producing high predictive accuracy means more crimes can be intersected before they actually occur

$$\text{Pred. Acc.} = \frac{no\ of\ crimes\ captured}{Total\ no\ of\ crime}$$

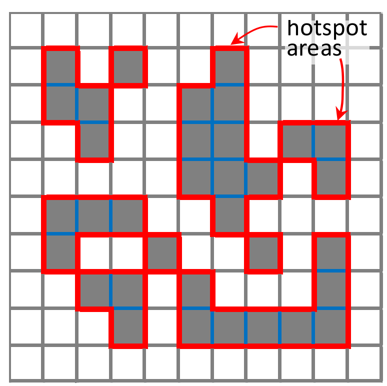
Evaluation metrics

2. **Compactness Index (CI)**

measures the ease at which a defined hotspot can be patrolled



Map A



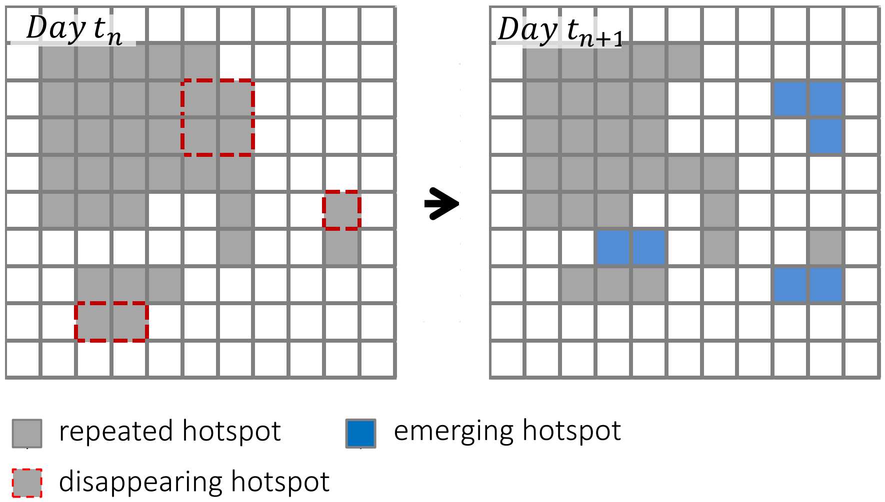
Map B

✓ From police standpoint, Map A is easier to patrol than Map B because of higher connectivity ratio of hotspot units

Evaluation metrics

3. Dynamic Variability Index (DVI)

measures the extent to which the predicted locations change between consecutive predictions

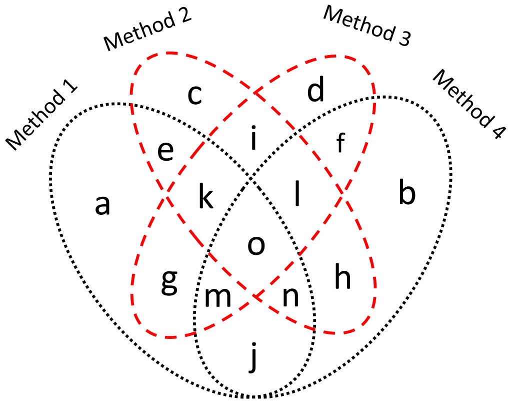


- ✓ DVI helps to distinguish between different methods based on the type of crime risk they detect e.g. short-term risk, long-term risks.

Evaluation metrics

4. **Complementarity**

measures the extent to which different methods detect the same and/or different crimes



Venn diagram showing complementarity of four different methods; a – o indicate number of crimes

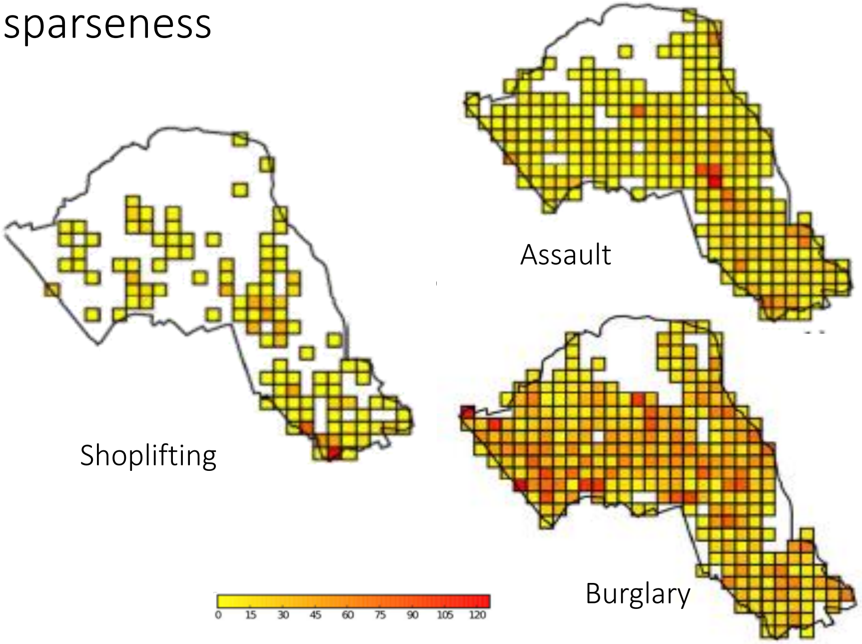
✓ helps to reveal how much improvement a method made relative to other methods.

❖ Adepeju M., Rosser G. & Cheng T. (2016) Novel evaluation metrics for sparse spatiotemporal point process – a crime case study STPP. Intl. J. Geog. Info. Sys. 1(22)

Case Study - London Borough of Camden

Aim: To demonstrate the utility of the proposed evaluation framework

Dataset: 3 crime types of varied level of ST sparseness



Predict-evaluate routine:

- ✓ For each method, generate hotspots at day tn,
- ✓ Evaluate the hotspot for one day ahead (tn+1)
- ✓ Repeat for 100 consecutive days

Results

mean hit rate, mean CI and mean DVI

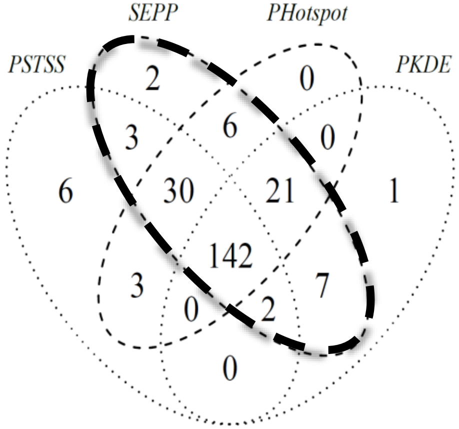
Crime Type	Method	Accuracy	Hotspot compactness	Variability
		Mean Hit rate	Mean CI	Mean DVI
Shoplift	PSTSS	81.3	0.42	14.9
	PKDE	74.3	**0.55**	2.7
	SEPP	**91.5**	0.31	6.0
	PHotspot	85.1	0.37	**19.2**
Violence	PSTSS	46.5	0.46	10.8
	PKDE	51.7	**0.54**	2.6
	SEPP	**59.7**	0.12	4.5
	PHotspot	52.2	0.32	**21.1**
Burglary	PSTSS	34.4	**0.51**	3.7
	PKDE	38.8	0.50	2.3
	SEPP	**47.4**	0.02	1.4
	PHotspot	34.9	0.30	**5.3**

Evaluation metrics of Camden crime prediction at 20% coverage level.

Results

Complementarity

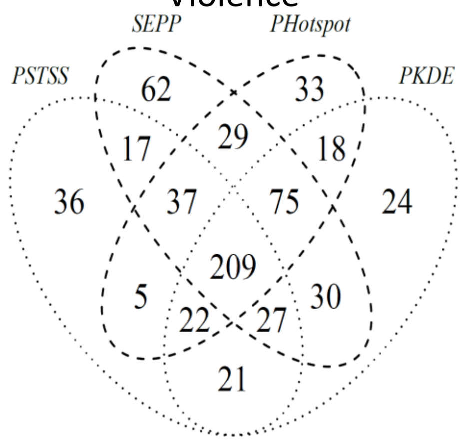
Shoplifting



Undetected: 9

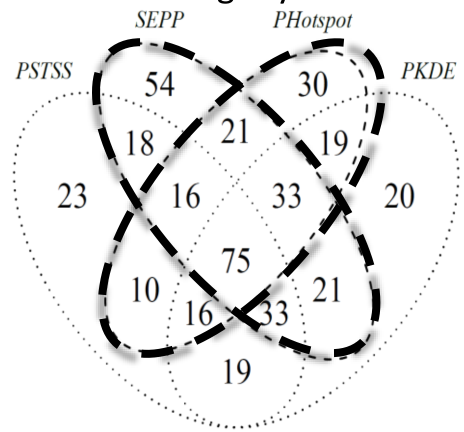
- ✓ Total = 223
- ✓ SEPP captures 213

Violence



Undetected: 155

Burglary



Undetected: 165

- ✓ Total = 526
- ✓ PHotspot captures 220
- ✓ SEPP captures 271

Venn diagram showing the total number of crimes identified by each method at a fixed coverage of 20% in Camden

Discussions

- Trade-offs:
  - ✓ Predictive accuracy vs. Compactness (ease of patrol)

- DVI reveals that certain methods are best suited for specific type of risk patterns. For example, PSTSS capture emerging risks patterns while PKDE captures persistent risk patterns

- Complementarity suggests that results of some methods can be combined for better performance - ensemble predictions.