This is a pre-print of a commentary accepted for publication by the Security Journal in 2008. It is a pre-copy-edit version of the letter that is found at http://www.palgrave-journals.com/sj/journal/v21/n4/abs/sj20087a.html

Response to Levine 2008, 'The hottest part of a hotspot: comments on "Spencer Chainey, Lisa Tompson and Sebastian Uhlig, The utility of hotspot mapping for predicting spatial patterns of crime'. Security Journal.

Spencer Chainey[1], Lisa Tompson[2], and Sebastian Uhlig[3]

[1] Director of Geographical Information Science, Jill Dando Institute of Crime Science, University College London, Brook House, 2-16 Torrington Place London. WC1E 7HN
[2] Research Fellow, Jill Dando Institute of Crime Science, University College London, Brook House, 2-16 Torrington Place London. WC1E 7HN
[3] Project Consultant, Informed Solutions, The Old Bank, Old Market Place, Altrincham, Cheshire. WA14 4PA, UK

Levine (2008) provides a useful contribution to the assessment of hotspot techniques in their utility for predicting spatial patterns of crime – considering in particular the application of the nearest neighbour hierarchical clustering technique (Nnh) and convex hulls. However, we think that there are some inaccuracies in his statements, and errors and lack of rigour in his methodology that make his results misleading and undermine his conclusions.

Levine (2008) suggests that "the methodology they [Chainey, Tompson and Uhlig, 2008] adopted to compare several hotspot techniques was limited and, consequently, raised doubts about the conclusions they drew". We argue that the methodology is not limited (indeed Levine (2008) refers to it as a 'good criterion for comparing different hotspot techniques'). Instead, we do acknowledge that the range of hotspot mapping techniques that can be applied against the methodology can be expanded. Our purpose, as described in Chainey et al (2008) was to concentrate on exploring the predication accuracy of the most common hotspot mapping techniques that are applied by practitioners and as evidenced by Weir and Bangs (2007) in their survey of crime analysts. Our assessment was that the thematic mapping of small geographical units (e.g. Census Output Areas), spatial ellipses produced by the STAC routine, grid thematic mapping, and kernel density estimation were the most common hotspot mapping techniques used in practice. We also illustrated this with reference to several examples of their application. We could have considered many others but felt that the four mapping techniques we analysed would offer the most significant contribution to researchers and practitioners. Indeed, we even recommended that other researchers apply the Prediction Accuracy Index (PAI) (Chainey et al, 2008) to other techniques so that other contributions could be made to this research theme. Levine's contribution in testing the Nnh technique and convex hulls is therefore very welcomed. Hence, we argue that our conclusions are sound, and offer a strong foundation for other researchers to apply the PAI to other mapping techniques, compare against our results and help to develop research into the prediction utility of hotspot mapping.

We also believe that there are a number of inaccuracies in Levine's methodology, which in turn have produced some misleading results. These are in relation to the parameters that Levine (2008) uses in his analysis and his sample:

Parameters used by Levine (2008) for KDE, STAC ellipses and Nnh

We suggested (in Chainey et al, 2008) that the parameters used for hotspot mapping techniques should be based on the data that is being analysed, rather than on fixed settings that are applied to all different types of data. In our research we used Hotspot Detective (Ratcliffe, 2002) to generate KDE surfaces. The parameters that Hotspot Detective determines as default settings are specific and are chosen after an analysis of the spatial characteristics of the input data. We then applied these settings to other techniques that used the same corresponding input data (i.e. data for the same input data period, for the same measurement date and for the same crime type).

Cell size values were calculated by dividing the shorter side of the minimum bounding rectangle around the crime input data by 150 (Ratcliffe, 2004). In Hotspot Detective the calculation of the default bandwidth value used for generating KDE surfaces is not divulged to users. It is known, however, to be a function of the shorter side of the minimum bounding rectangle surrounding the crime data, divided by a number that provides a suitable value without requiring a significant number of iterations to generate a representative KDE surface (Ratcliffe – personal communication). For example, the parameters determined for input data of residential burglary that covered an input data period of 3 months were a cell size value of 48m and a bandwidth value of 240m. For theft of vehicle data that covered an input data period of 1 week the number of records and their spatial distribution would be different and hence require different parameter settings. For these data the cell size value was 37m and the bandwidth value was 185m (see table 1 for a full listing of parameter values that illustrate different settings are required for different data). Levine used a cell size of 49m and 245m on all his tests of KDE, STAC and Nnh for data from Houston, Texas. These were based on parameter settings that we determined for three of the eighty different input datasets in our research using crime data for Camden and Islington in London. Levine should have determined the specific parameter settings (such as the KDE cell size and bandwidth) from a similar analysis of his input data. Instead he uses the parameters we determined which are not comparable to his dataset: for example, 4243 theft of vehicle offences in Camden and Islington in comparison to 21352 offences of burglary of vehicle offences for the much larger geographic area of Houston. We suggest that this error in his methodology undermines his conclusions and renders his results as misleading.

PAI values for different input and measurement date periods

Levine (2008) performed analysis that compared 12 months of input data from 2005 to 12 months of measurement from 2006. We performed our analysis for a combination of 100 different input data and measurement data periods for four different crime types, for two different measurement dates (i.e. 800

experiments) in order to provide a representative sample of PAI results. We suggest that for Levine to determine that certain methods are better than others requires a similar rigour to that which we applied. A sample that is based on analysis of only one input data period and one measurement period for one measurement date may produce results that are not fully representative of the performance of hotspot mapping techniques.

We suggest that Levine repeats his analysis following the methodology we have described in Chainey et al (2008) if he wishes to confirm the conclusions he arrives at. We do consider his explanation of the Nnh technique as being particularly worthwhile and that the application of convex hulls for determining crime hotspots as offering a potentially useful technique for predicting spatial patterns of crime. Similarly, Chainey and Tompson are currently researching the application of the Nnh technique, convex hulls and the Gi* statistic (Getis and Ord, 1996) for predicting spatial patterns of crime using the dataset for Camden and Islington. We plan to publish the results at some future point.

Hotspot Detective default values for all crime types and input data time periods (cell size / bandwidth) for 01/01/03

Input data time period	Residential Burglary		Street Crime		Theft from vehicle		Theft of vehicle	
	C.S.	B.	C.S.	B.	C.S.	B.	C.S.	B.
12 months	49	245	49	245	49	245	48	240
6 months	49	245	49	245	49	245	46	230
3 months	48	240	48	240	49	245	46	230
2 months	48	240	48	240	48	250	46	230
1 month	46	230	46	230	46	230	46	230
2 weeks	44	220	45	225	45	225	45	225
1 week	44	220	44	220	45	225	37	185
3 days	38	190	44	220	44	220	34	170
2 days	38	190	31	155	40	200	34	170
1 day	38	190	2	10	40	200	27	135

(a)

Hotspot Detective default values for all crime types and input data time periods (cell size / bandwidth) for 13/03/03

Input data time period	Residential Burglary		Street Crime		Theft from vehicle		Theft of vehicle	
	C.S.	B.	C.S.	B.	C.S.	B.	C.S.	B.
12 months	46	230	46	230	46	230	45	225
6 months	45	225	46	230	46	230	45	225
3 months	45	225	44	220	45	225	45	225
2 months	45	225	43	215	45	225	45	225
1 month	43	215	43	215	44	220	44	220
2 weeks	41	205	43	215	42	210	42	210
1 week	39	195	42	210	41	205	42	210
3 days	39	195	39	195	40	200	36	180
2 days	30	150	39	195	40	200	31	155

1 day	25	125	38	190	32	160	27	135

(b)

Table 1. Hotspot Detective KDE default values (C.S. – cell size and B. – bandwidth) for each crime type and each period of input data, for (a) a measurement date of the 1st January 2003 and (b) a measurement date of the 13th March 2003. Bandwidth measures were used to determine alternate search radii for spatial ellipses, and alternate grid thematic mapping cell sizes. For example, a grid thematic map was generated from one month of street crime data, for the 01/01/2003 measurement date, using a grid cell size of 230m. Reproduced from Chainey et al (2008).

References

Chainey,S.P., Tompson,L., Uhlig,S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. Security Journal.

Getis, A. and Ord, J.K. (1996). Local spatial statistics : an overview. In Longley, P. and Batty, M. (Eds), Spatial Analysis: Modelling in a GIS Environment. GeoInformation International: Cambridge, England, 261-277.

Levine, N. (2008) 'The hottest part of a hotspot: comments on "Spencer Chainey, Lisa Tompson and Sebastian Uhlig, The utility of hotspot mapping for predicting spatial patterns of crime'. Security Journal.

Ratcliffe, J., (2002) *HotSpot Detective 2.0 for MapInfo Professional 7.0*. Available online at http://jratcliffe.net/hsd/

Ratcliffe, J. (2004) *HotSpot Detective for MapInfo Helpfile Version 2.0*