**SPACE-TIME ANALYSIS OF CRIME PATTERNS IN CENTRAL LONDON**

T. Cheng*, **D. Williams**

*Dept. of Geomatic Engineering, University College London, Gower Street, London, WC1E 6BT UK
tao.cheng@ucl.ac.uk

Commission II, WG II/3,

**KEY WORDS:** Exploratory space-time analysis, visualization, crime analysis, data mining, space-time cube, spatial scan statistics, self-organizing map, multi-resolution

**ABSTRACT:**
Crime continues to cast a shadow over citizen well-being in big cities today, while also imposing huge economic and social costs. Timely understanding of how criminality emerges and how crime patterns evolve is crucial to anticipating crime, dealing with it when it occurs and developing public confidence in the police service. Every day, about 10,000 crime incidents are reported by citizens, recorded and geo-referenced in the London Metropolitan Police Service Computer Aided Dispatch (CAD) database. The unique nature of this dataset allows the patterns to be explored at particularly fine temporal granularity and at multiple spatial resolutions. This study provides a framework for the exploratory spatio-temporal analysis of crime patterns that combines visual inquiry tools (interactive animations, space-time cubes and map matrices) with cluster analysis (spatial-temporal scan statistics and the self-organizing map). This framework is tested on the CAD dataset for the London Borough of Camden in March 2010. Patterns of crime through space and time are discovered and the clustering methods were evaluated on their ability to facilitate the discovery and interpretation of these patterns.

1. **INTRODUCTION**

Crime and criminality impose large economic and social costs. Citizens, particularly in big cities, are negatively affected. In the past, crime analysis has suffered from abstracting the spatial aspect of crime from its temporal aspects. Better understanding of the distribution of crime incidents over time, termed the temporal signature of crime (Ratcliffe 2010), has potential to further the understanding of crime patterns and the prevention of crime and criminality in general (Ratcliffe 2010; Nakaya and Yano 2010; Brunsdon et al. 2007; Roth, et al. 2010). Spatio-Temporal Data Mining is the extraction of latent knowledge, relationships or patterns from spatial-temporal datasets (Yao 2003). By extension, Exploratory Spatio-Temporal Data Mining (ESTDM) is the extraction of latent knowledge, relationships or patterns from spatial-temporal datasets through analysis not governed by a particular hypothesis or aim. ESTDA is inherently challenging due to the complex nature of spatio-temporal data, data types and structures as well as the lack of efficient methods for exploring it (Yao 2003).

Geographic Visualization or Geovisualization (GeoViz) is the branch of visualization that applied to every problem solving step of geospatial analysis - from hypothesis generation to knowledge discovery to the final presentation (Gibin, et al. 2008). Exploratory Spatio-Temporal Data Analysis (ESTDA) coupled with geovisualization tools has the potential to facilitate the exploration of crime patterns required to create intelligence and improving the efficiency of policing. This study intends to investigate a framework for the exploratory spatio-temporal analysis of crime patterns which combines geovisualization tools with cluster analysis. This framework will then be applied to a complex, real-time CAD police dataset of crime incidents for the London Borough of Camden.

2. **METHODOLOGY**

ESTDA and Geovisualization shall be applied through a general methodology which features a visual inquiry tools and cluster detection. The methods and visualization tools used in this study may be seen in figure 1 and are categorized according to their use for purely spatial, purely temporal or spatio-temporal analysis.

![Figure 1. Visual inquiry and cluster detection tools to be used in this study.](image)

A progressive deepening approach to the STDM (Han 2003) shall be applied to ESTDA in this context, to allow analysis through the continuum from micro to macro scales: spatially - from a crime point cloud to crime aggregated to Lower Super Output Area (LSOA) geography and temporally – from crime incidents per second, to hourly, daily, weekly and monthly temporal aggregation. This methodology can be applied to different datasets and processes by changing the tools and the spatial and temporal resolutions used.

* Corresponding author.
2.1 ESTDA - Visual Inquiry

Cartograms, scatter plots, Chloropleth maps, histograms, parallel coordinate plots, area graphs, a custom animation and a 2.5/3D scatter plot in a space-time cube were shall be used at varying spatial and temporal resolutions.

2.1.1 Animation: The potential of an animation to illuminate spatio-temporal trends has been widely noted in research (Brunsdon et al. 2007; Li 2005; Ratcliffe, 2010). Animations are relatively underutilized in crime analysis (Brunsdon et al. 2007) due to lack of sophisticated, easy to use products and training and tools to produce them (Ratcliffe, 2010). The open source code library published by Roth, et al. (2010) will be employed in this project to create a custom animation with advanced features such as the interactive temporal histogram in addition to the requisite VCR controls (R. E. Roth 2009). This code library extends the Google Flash API and was formed from the code used to make the GEOVISTA CrimeViz application.

2.1.2 STC Scatter Plot: Animations and scatter plots in space-time cubes (Kristensson, et al. 2009) are suitable for representing point clouds in time and space. Scatter plot in space-time cubes are suitable for small point clouds, <2000 points, as over plotting degrades the display (Klemelä 2009). A STC scatterplot will be used here; bar lines will be added and colour will be manipulated for clarity.

2.2 ESTDA - Cluster Analysis

The planar and network kernel density estimation (KDE), spatio-temporal scan statistics (STSS) and self-organised map (SOM) technique will be used for cluster analysis.

2.2.1 Kernel Density Estimation: Kernel Density Estimation is a commonly applied interpolation and hotspot technique (Hagenaier et al. 2011). A statistical function, based on an empirical probability function by Silverman (1986), is used to convert a point set in geographic space into a density surface. KDE has been cited in literature as exaggerating high density areas and the results of KDE are well known to be highly dependent upon the values selected for the cell size and bandwidth parameters (Nakaya and Yano 2010). Setting the bandwidth equal to the average distance between the kth nearest neighbour is common in crime science practice (Nakaya and Yano 2010). While k=20 is often chosen for conventional 2D mapping of city wide crime (Harada and Shimada 2000), the choosing of k is somewhat subjective. A larger k value provides a smoother density surface which is suitable for analysis over a wide area whereas a smaller kernel bandwidth can identify hotspots at specific street corners (Chainey and Ratcliffe 2005). A heuristic for choosing the cell size parameter is set the value to the result of dividing the length of the shortest side of the minimum bounding rectangle of the study area by 150.

KDE has been extended to analyse crime rates by an isosurface approach (Brunsdon et al. 2007) or a volume rendering approach (Nakaya and Yano 2010) may be taken to approximately represent it.

Crime incidents are affected and constrained by the built environment, road networks, alleyways within buildings etc (Meyer 2010). Spatial network analysis is based on the proposition that because of this constraint, analysis in a strictly standard planar sense is likely to lead to false conclusions (Okabe et al. 2006). Planar and network analysis share basically the same model of a network located in a plane with incidents represented as points but for network analysis the distance between points is taken as the shortest path distance along the network.

Research is required to determine whether the use of network spatial analysis provides any significant benefit over planar spatial analysis (Meyer 2010). Yamada and Thrill (2004) found an overall benefit to using the network K Function over the planar K Function which they found exaggerated accident clusters. While, Spooner et al. (2004) found the combination of network spatial analysis with a planar kernel density for visualization of the network patterns extremely informative, the planar kernel exaggerated the importance of the patterns which the network K Function found statistically insignificant. Nakaya and Yano (2010) have identified the need to combine network spatial analysis and statistics with the space-time cube especially in focused analysis of crime patterns in small geographic areas (Nakaya and Yano 2010).

2.2.2 Spatial Scan Statistics: Spatial Scan Statistics may be used to determine the location of spatio-temporal hotspots. The crime incident positions and the times at which they occurred are input into the model. As the analysis algorithm is initiated, the cylindrical scan window moves throughout the study areas and continuously changes its dimensions: base radius refers to the spatial dimension and height controls the temporal dimension. The result is a multitude of overlapping cylinders of various sizes. The number of crimes actually observed is compared to the number of crimes that would be expected from the calculations of a Poisson distribution model or a space-time permutation model. A hotspot or cluster is a cylinder whose number of observed crimes is significantly larger, statistically, to the expected value. Statistical significance in this case, is determined by a Monte Carlo simulation of 999 repetitions. The maximum percentage of population at risk is the most subjective criteria to be specified. Most users use the default value of 50% as it is recommended by the user guide and because SatScan™ automatically searches for small and large clusters (Pfeiffer, et al. 2008). Boscoe et al noted that the most significant cluster is often times a large, relatively low risk cluster with low prevalence (Pfeiffer, et al. 2008). Nakaya and Yano (2010) presented the clusters in three dimensions within a space-time cube by representing them as space-time cylinders. These cylinders were created in the ESRI ArcScene™ 3D Extension.

2.2.3 Self Organizing Maps: The Self Organizing Map (SOM) is an unsupervised learning neural network (Hagenaier et al. 2011). It allows simple, geometric relationships to be produced from vector quantization analysis of complex multi-dimensional datasets (Kohonen, et al. 1996). It facilitates multivariate analysis through a combination of clustering and dimension reduction (Guo, et al. 2006) while preserving topology (Kohonen, et al. 1996).

The recursive, regressive algorithm conducts several training phases during which the units move from being roughly configured to being finely tuned (Andrienko et al. 2010). The SOM method assumes a two dimensional rectangular or hexagonal grid (Kohonen et al. 1996) and therefore produces a 2D grid of the output data which may be used to create static visualizations (Andrienko et al. 2010) such as U, distance or similarity matrices, 3D SOM surfaces and SOM component plans (Koua et al. 2006).
The SOM method is an improvement over spatial scan statistics and STKDE because it allows the attributes of the geographic space to be modelled along with the spatio-temporal data and was used to assess spatio-temporal crime patterns in Houston, Texas (Hagenaier et al. 2011). While Hagenaier et al. used distance matrices presented as a series of interactive linked views, one for each time stamp, Andreinko et al. (2010) used two SOM groups (one for spatial distributions and one for temporal variations) and combined these linked views with interactive visualization tools such as time series graphs, time arrangers and periodic pattern displays. Colour-coding was used to allow the viewer to connect data across different visualizations and allowed the patterns to be recognized in both time and space. An approach similar to Andreinko et al. (2010) shall be taken using the open source program VIS-STAMP (Guo et al. 2006).

3. CASE STUDY

3.1 Study Area & Dataset

The methodology was applied to data from the borough of Camden, London, UK. Camden is a central, inner London borough which is approximately 22 kilometres squared in area. A dataset of nine thousand five hundred and ninety-two (9592) geo-coded incident reports, recorded for March 2010, from the Computer Aided Dispatch (CAD) System of the Camden Borough Police was obtained. The unedited dataset contained over 100 distinct classes of offences. A list of class codes and titles were used to identify the relevant offences such as ‘violence against person’, ‘sexual offence’, ‘burglary dwelling’, ‘robbery’, etc. Not all calls that are made result in an incident being “crimed” i.e. a CRIS number being assigned and the incident being featured on a crime map as all incidents do not lead to criminal prosecution (Ratcliffe and McCullagh 1999). The details of what constitutes criminal activity which warrants an incident being “crimed” is not within the domain of this study; the incidents assigned a CRIS number were identified as crimes and were isolated for further investigation. Two thousand two hundred and ninety (2290) criminal offences were recorded; broken down by type in the table below. The most frequently occurring offence was “Other Theft” followed by “Malicious Nuisance Calls”. Traffic violations were the least commonly occurring offence over the study period.

Although police call out data is plagued by over and under reporting (Sherman et al. 1989) it is still a useful indicator of criminal activity (Brunsdon et al. 2007). This confidential dataset has extremely fine temporal granularity as the date and time to the nearest second are recorded, while the most detailed spatial granularity available; the crime points have been aggregated to a 250m x 250m grid and that the southern regions, particularly Holborn and Camden had more crime incidents than the northern region. The navigation of the 3D cube is not intuitive. However, it provided a succinct overall picture which allowed a few spatial abnormalities to be found.

3.2 Visual Inquiry

The visual inquiry was conducted on data aggregated to the entire month to get a general overview. Initial visual inspections of the data aggregated for the entire month revealed clustering in space, time and in together space-time. A scatter plot of the incident coordinates was used to assess the spatial distribution of the incidents. Crime incidents were present throughout the entire study region but potential clusters were highlighted near Holborn, Camden Town and south Swiss Cottage.

3.2.1 Spatial: Histograms, 2D scatter plots, cartograms, parallel coordinate plots and box maps indicated the presence of spatial clustering of high crime counts outliers in the Holborn and Camden regions. The highest crime rate, or 101.18 incidents per 1000 persons, was found in LSOA 28D in Holborn. These visualizations were simple to create and easy to interpret, with the exception of the scatter plot which suffered from over plotting. While the over plotting made highlighted the clusters, it marred the aesthetic of the diagram (not shown).

3.2.2 Temporal: An area graph (figure 2) of the incidents against the time of occurrence revealed that crime was more likely to occur between the hours of 9 am and 9pm, with 2 pm and 5pm being the most likely times. Crime incidents were least likely to occur from 3am to 7am. This holds to the tenants of crime pattern, routine activity and temporal constraint theories – crime occurs during the regular activities of life whenever the right combination of offenders and targets occurs in space.

The temporal variation of thefts as opposed to all crime types was very similar except the thefts had only one peak, at around 3pm, and a longer trough from 2am to 10am. This warrants further investigation. This visualization was simple to create and easy to interpret. However, some manual processing was required to create hourly and daily counts.

Spatio-Temporal: The interactive animation showed crime incidents dispersed over the study area. The animation was intuitive to use and relatively simple to build. It could be made more useful by including a search feature and by allowing different levels of temporal aggregations. No distinct trends were found, but a few stable crime spots were identified. The space-time cube scatter plot allowed general trends to be seen more clearly. The incidents of each week were symbolized in a different colour. It was apparent that the points were aggregated to a 250m x 250m grid and that the southern regions, particularly Holborn and Camden had more crime incidents than the northern region. The navigation of the 3D cube is not intuitive. However, it provided a succinct overall picture which allowed a few spatial abnormalities to be found.
3.3 Cluster Analysis

The Global and Local Moran’s I tests of global spatial autocorrelation indicated a moderate positive autocorrelation for all crime incidents and for the largest occurring crime type “Other Theft”. This suggests a clustering of high crime values in space for all crimes. These values was tested and deemed to be statistically significant after 999 permutations. The results agreed to three significant figures when either the queen’s case or rook’s case contiguity models were used. The Local Moran’s I statistic identified high-high clusters were identified in parts of Bloomsbury, Holborn & Covent Garden and Camden Town and low-low clusters were identified in LSOAs 4B, 2C, 5D.

Planar and Network KDE revealed hotspots in the vicinity of Holborn and Camden Town. KDE was conducted using the bandwidth and cell size parameters chosen by Ratcliffe and Chai ney in a case study of ‘theft of vehicle’ incidents for the borough of Camden (Chai ney and Ratcliffe 2005) to ensure that this KDE application would be comparable to theirs. The Bloomsbury, Holborn & Covent Garden regions feature seven intense hotspots and thirteen supplementary hotspots in a dispersed pattern throughout the neighbourhood. In stark contrast, the Camden Town features two hotpoints: a primary hotpoint in the vicinity of the Camden Town tube station and a secondary hotpoint in the vicinity of the popular shopping centre Camden Lock.

Both the planar and network KDE were useful in identifying the clusters though the network KDE algorithm appeared to be less dependent on input parameters. A prototype wall map was created which demonstrated that such maps could be very useful in displaying spatio-temporal crime along the road network although potential interpretability problems may arise with more a complex road network.

3.3.1 Spatial Scan Statistics: SatScan™ was used to identify statistically significant clusters in space and time. First, the raw point cloud was analyzed using the space-time permutation model. Then, a space-time Poisson model was used on the points aggregated to the population centre of each LSOA from the 2002 census data being used as an estimate of the population at risk. A clusters were deemed statistically significant when \( p \leq 0.01 \) for both methods. The space time permutation algorithm found no significant clusters for the entire period when tested against a significance level of 1% i.e. \( p=0.05 \). The test was repeated while varying the maximum spatial cluster size, maximum temporal window size and the spatial cluster shape parameters. The \( p \) values responded to the changes in parameters but no change resulted in a statistically significant cluster. The clusters that were found, albeit insignificant, followed the finding of who reported the space-time permutation model producing large, low risk, statistically significant clusters (Pfeiffer, et al. 2008). However, the space-time Poisson model was more useful and provided two highly significant (\( p \approx 0 \)) clusters.

As expected the clusters were found in the Holborn/ Bloomsbury area and the Camden Town Area. Both clusters had duration of 14 days but the start and end dates differed. The Camden area cluster had a smaller radius and a greater relative risk, than the wider cluster in the Holborn, Bloomsbury area. This corroborates the earlier picture of Camden Town being a hotspot, and the Holborn, Bloomsbury Covent Garden area being a comprised of multiple hotspots. Alternate views of the hotspots can be seen in figures 3 and 4.

Figure 3. Profile view of Crime cluster cylinders visualized in a Space-Time Cube. Each bar line delineates one week, orange represents the Holborn cluster and red the Camden cluster.

This visualization technique is extremely effective and useful with the results being clear and easy to interpret. It may be improved by adding a labelling the bar lines and shadow beneath the cylinders so that when the map is rotated the user can still know exactly where the cluster is centred.

3.3.2 Self Organizing Map: The visual analytics mantra (Keim et al. 2006) was applied to analyze the temporal variation of patterns of crime in space (Space-in-Time) and the distribution of the temporal patterns in space (Time-in-Space) (Andrienko et al. 2010) using open source software package VIS-STAMP (Guo, et al. 2006). This package implements the Self-Organizing Map method to produce a customized U Distance matrix which it pairs via interactive, linked plots to a Parallel Coordinate plot, which also acts as a legend, a matrix of maps for each study period and a reorderable matrix which categorizes each area under display into a cluster for the periods under investigation. The SOM method was computed using a network of 7x7 nodes for a dataset of crime rates for 133 LSOAs for all 17 crime types for each of the 4 weeks of the study period.

For the Space-In-Time Analysis the training of the 7x7 matrix produced 48 clusters of multivariate crime profiles. The VIS-STAMP program displays them using a U-Matrix (fig 43) with a traditional hexagonal layout but with the clusters represented as coloured circles. The size of the circle is linearly proportional to the number of data items it is comprised of while the greyscale of the nodes represents the multivariate dissimilarity between the node and its neighbour. There is the greatest dissimilarity between blue clusters, e.g. 43 and 42 or 38, while the large purple and orange clusters are relatively similar to their neighbours e.g. 48 and 41 or 11 and 6. The most common cluster, 11, (the largest coloured circle in the U Matrix and the thickest line on the PCP) featured LSOAs with a crime rate of <1 “other thefts” and rates of <0.2 violence against a person and <0.1 “shoplifting” per 1000 persons.
The map matrix is useful to allow the spatial variation to be examined with the topology of the clusters in geographic space retained. The map matrix revealed that approximately 17% of the LSOAs had no crimes reported for at least one week in the study period. Additionally, apart from the LSOA’s in the Hampstead area which were predominately in the blue or green clusters there were no strong trends in spatial variation in the clusters.

For the “Time-in-Space” analysis thirty-eight (38) temporal profile clusters are discovered with the medium blue cluster (33) dominating the visualization. This cluster represents a constant temporal profile of value approximately equal to zero i.e. no change in crime rate for the study period and a crime rate of zero for a particular crime type for the study period. Therefore, for this study it is very common for no crimes to be reported in an LSOA for one specific crime for the entire study period. There was at least one crime reported for all LSOA’s for the study period. Apart from the prevalence of the blue cluster, the crime type with the most striking trend is “Other Thefts”. The space-variable matrix shows a strong grouping of LSOA’s belonging to the dark red cluster (7) with 15% of LSOA’s having a similar temporal variation for the “Other Thefts” crime type.

The SOM method is extremely effective and very useful particularly for the visualization and analysis of the attribute space. The clusters are clear and distinct and with the linked views of the PCP, reorderable matrix and map matrix, the trends can be clearly interpreted.

4. CONCLUSIONS

Analysis of crime patterns in time and space is very complex. The use of a Police CAD dataset made this study particularly complex because each crime incident was referenced to a small temporal granularity (seconds), a small spatial unit (a point) and had a multidimensional attribute space. The use of a CAD dataset for the study of crime patterns in the UK was not found in Literature. This paper shows the use of numerous techniques for visualizing crime incident patterns at varying, complimentary, spatial and temporal scales. Distinct visualizations support different tasks and the representation of data at certain spatial and temporal scales. Spatial scan statistics best supports the identification and description of spatio-temporal clusters when the attribute space need not be studied/ visualized in great detail. It is suitable at all levels of spatial and temporal aggregation. The results of this method are effectively and usefully displayed using space-time cylinders in a space-time cube. The self organizing map method may be used at varying spatial and temporal scales for spatio-temporal analysis of crime patterns particularly where detailed exploration of the crime attributes or the socio-economic context of the crime patterns is required. While the SOM package used here (VIS-STAMP) (Guo, et al. 2006), does not support the analysis of raw point data, the SOM-PAK package (Kohonen, et al. 1996) for Matlab may be used for that purpose.

ACKNOWLEDGEMENTS

This research is part of the CPC Project (Crime, Policing and Citizens) supported by UK ESPRC (EP/J004197/1).

REFERENCES


