

## Spatial Analysis of South African Crime Data

Elsabe Smit

University of the Witwatersrand, Johannesburg, South Africa

e-mail: [elsabe.smit@wits.ac.za](mailto:elsabe.smit@wits.ac.za)

### Abstract

Crime is a prominent issue in South Africa. The high crime rate has given the country a reputation as a crime capital. Although there has been a reduction in serious crimes over the last few years, the fight against crime is still a key priority of the Government. An understanding of the distribution of crime in space can aid in developing crime prevention strategies. This paper describes how principal component analysis can be used to derive general and specific measures of reported contact crime. The smoothed spatial distributions of these measures are mapped and evaluated for the period 2003 to 2011.

Keywords: Principal component analysis, crime mapping, spatial smoothing.

### 1. Introduction

The high prevalence of crime in South Africa (SA) is well-known throughout the world. Although the total crime levels in SA decreased by 21% between the 2002/03 and 2011/12 financial years, the incidence of violent crime is still very high. During 2011/12 the murder and violent crime rate per 100,000 population was 30.9 for SA, compared to 17 for Africa and 6.9 globally (Newham, Burger, Gould and Lancaster, 2012). South Africans' perceptions of crime reflect the fact that crime is a major issue in the country. The Victims of Crime Survey (VOCS) is a national survey aimed at obtaining information about the general population's perception of safety and security, as well as detail regarding personal and household victimisation. The VOCS was carried out five times by Statistics South Africa (Stats SA) and the Institute for Security Studies (ISS) from 1998 to 2012. According to the VOCS (Stats SA, 2012) approximately half of the households in the population physically protect their homes and 23.7% use some form of vehicle protection. A third of the population believed that there was an increase in violent crimes over the previous two years, while 38% felt that it decreased.

Holtman and Domingo-Swarts (2008), and Burger, Gould and Newham (2010) state that the SA government is committed to the fight against crime. Many new policies related to safety, security and crime prevention have surfaced between 1994 and 2005, such as the National Crime Prevention Strategy in 1996 and the Alliance for Crime Prevention in 2005. The latter is a collaboration between government and organisations like the Council for Scientific and Industrial Research and the ISS. In addition to these policies the South African Police Service (SAPS) budget increased drastically, to be used for more manpower, technology and training.

Burger *et al.* (2010) argue that the fight against crime in SA does not necessarily require more police officers but rather "*smarter and more accountable policing*". According to Ratcliffe (2010) criminal justice agencies should become proactive in order to prevent crime, and this can be facilitated through accurate prediction of potential crime locations using crime mapping techniques such as hotspot analysis. Crime hotspot analysis identifies areas where crime concentrations are higher

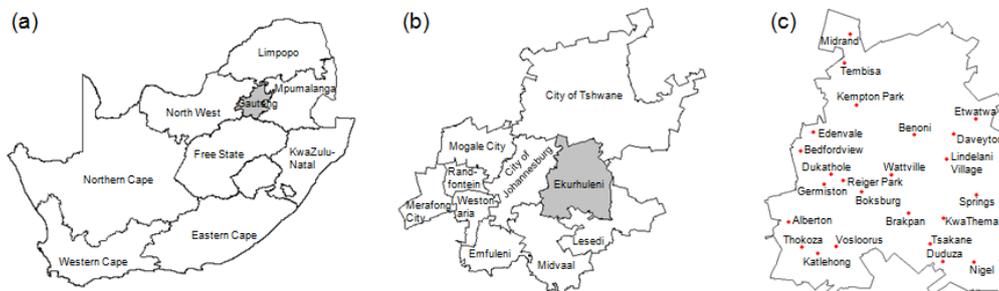
compared to other areas (Chainey, Tompson and Uhlig, 2008) and can be used to more effectively target police resources to areas of high crime intensity.

Newham *et al.* (2012) state that one has to look at different crime types to better understand the distribution of crime in SA. However, one criminal event can sometimes lead to another, for example a robbery can escalate to murder. This implies that various criminal activities do not necessarily occur in isolation. Principal component analysis (PCA) has been used in other research on different types of criminal activities to derive crime components (Ahaman, 1967; Salvati, Di Bartolomei, Rontos and Bisi, 2012).

The main objective of this research is to assess the use of PCA to derive general and specific components of crime, focusing on crimes against persons, also referred to as contact crimes. The distributions of these components are then evaluated for a particular area in SA, namely the Ekurhuleni Metropolitan Municipality (EMM) in an attempt to identify places of high risk or hotspots.

### 2. Study area

SA consists of nine provinces and each province is demarcated into municipal areas. Although Gauteng is geographically the smallest province, the 2011 census count of its population is 12.3 million out of a total of 51.8 million, making it the most populous province. Gauteng is also the financial and commercial centre of SA. EMM is one of ten municipal areas in Gauteng and covers an area of approximately 11% of the province (1924 square kilometres). The 2001 census count for the EMM was approximately 2.5 million. This increased to 3.2 million in 2011. There are 27 main places (cities and towns) within the EMM with Germiston as its capital. *Figure 1* shows the SA provinces, municipalities in the Gauteng province, and main place location for medium to large towns in the EMM.



**Figure 1:** (a) Provinces of South Africa, (b) Gauteng municipalities, and (c) Main places in Ekurhuleni Metropolitan Municipality

### 3. Data sources

The SAPS reported crime figures are published over time for each of the 1120 police stations in SA ([www.saps.gov.za](http://www.saps.gov.za)). There are 29 different crime categories, seven of which are classified as contact crimes. This includes murder, attempted murder, sexual crimes, common assault, aggravated assault with the intent to cause grievous bodily harm (GBH), common robbery, and aggravated robbery. For this analysis, the reported contact crimes for eight years covering the time period April 2003 to March 2011 were used. The geographic information, used to map the crime components and identify hotspots for the EMM is available from Stats SA ([www.statssa.gov.za](http://www.statssa.gov.za)), together with the population counts for cities, towns and suburbs within the municipality.

#### 4. Analysis

The SAPS count data for the seven contact crimes were linked to twenty main places in the EMM. Such a small sample size could produce erroneous results from a PCA. In an initial analysis, the PCA was evaluated for all main places in the whole of SA, Gauteng and EMM separately. This was done per year as well as aggregated over the entire time period. It was found that the resulting component structures for the respective PCA's were very similar. Although there were differences in the actual component loadings, there were very high correlations between the scores saved from each analysis ( $r > 0.95$ ). This indicated that the data are relatively stable over time and space. Subsequent analyses were thus performed on the aggregate counts for the seven crime categories over the entire time period and for the whole of SA.

The contact crime components were created through two different PCA's, namely an unrotated PCA on the observed crime count data, and a Varimax rotated PCA on standardised crime count data per main place. Since the number of crime events in an area is often related to population size it is expected that a PCA on observed count data will reflect this. In order to remove this population effect, the seven crime counts per main place were expressed as a percentage of the total crimes within the main place. The component loadings for the PCA's on the observed and standardised data are given in *Table 1*.

When using the observed count data, the first two principal components explain 89.2% of the variability. These components measure two unique dimensions of contact crime. The first component shows high positive loadings on all seven contact crime types and can be seen as a measure of Crime Intensity, while the second component depicts the contrast between misdemeanour and felony crime types, yielding a measure of Crime Severity.

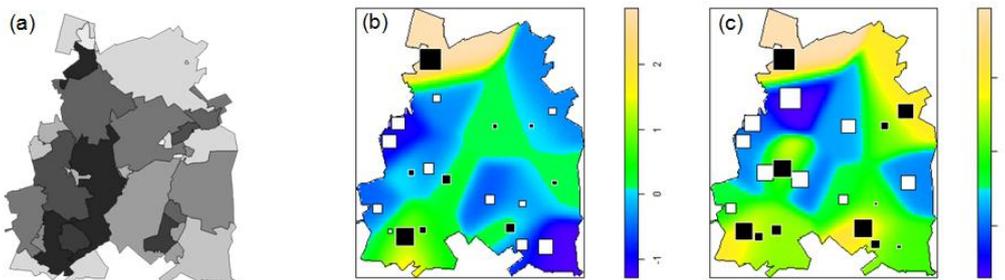
The PCA on standardised crime count data derives a set of three specific components of crime, with 79.8% of the variability explained. The Robbery Index is driven by both common and aggravated robbery, while the Contact Index describes the level of physical contact that victims may experience. Murder and attempted murder characterise the Murder Index.

	PCA on observed crime count data		PCA on standardised crime count data		
	Crime Intensity	Crime Severity	Robbery Index	Contact Index	Murder Index
% variance explained	80.1	9.1	39.3	26.6	13.9
Murder	<b>0.848</b>	<b>0.449</b>	<b>-0.138</b>	<b>0.415</b>	<b>0.720</b>
Sexual crimes	<b>0.959</b>	0.114	-0.349	<b>0.664</b>	0.179
Attempted murder	<b>0.904</b>	0.248	0.067	0.040	<b>0.895</b>
Assault with GBH	<b>0.904</b>	0.149	<b>-0.676</b>	<b>0.624</b>	-0.116
Common assault	<b>0.903</b>	-0.212	-0.229	<b>-0.906</b>	-0.242
Common robbery	<b>0.863</b>	<b>-0.440</b>	<b>0.876</b>	0.035	-0.221
Aggravated robbery	<b>0.878</b>	<b>-0.317</b>	<b>0.915</b>	-0.083	0.139

**Table 1:** Principal component loadings for PCA on observed and standardised crime count data

The five PCA scores are computed for all police stations in SA and evaluated for the EMM geographic area. In order to show the relative comparison of intensity of different components of crime within the EMM, these scores are standardised based on the ranges of values specific to the EMM area. A smoothed map is then created for each crime component over the area through Gaussian kernel smoothing with an optimal bandwidth choice for the kernel estimate. This is done using the functions *smooth.ppp* and *bw.diggle* in the **spatstat** library of **R**. The resulting maps indicate the unique distributions for the general and specific components of crime, identifying hotspot areas.

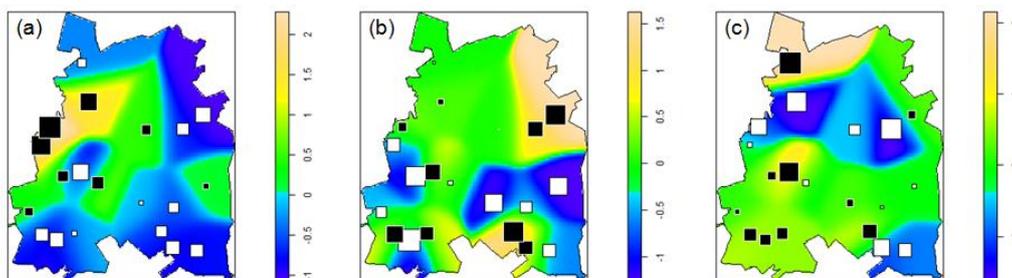
The smoothed map of Crime Intensity is given in *Figure 2b* and strongly relates to the distribution of population density in *Figure 2a*. As is often the case, the first PCA on observed data generally reflects some measure of size (Jolliffe, 2002). This is a very general component that indicates the “size of crime” in the area. In contrast to this the second component, namely Crime Severity, is more specific and displays a very different distribution in the EMM (*Figure 2c*).



**Figure 2:** (a) EMM population density, (b) Distribution of Crime Intensity, and (c) Distribution of Crime Severity

Further specific components of crime are found through the PCA on standardised crime counts and are give in *Figure 3*. To a certain extent, the Robbery Index (*Figure 3a*) is inversely related to Crime Severity (*Figure 2c*). It appears that robberies are concentrated in areas that are not generally affected by serious criminal activities. This could be a function of the demographic and social distributions in the area. On average the main places within these robbery hotspot areas are typically higher income neighbourhoods, which would logically attract burglars.

*Figures 3b* and *3c* show the unique distributions of the Contact Index and the Murder Index respectively. The eastern part of the EMM experiences a high intensity of serious physical contact crimes such as sexual crimes and assault with GBH. Murder appears to be dominant in the northern parts of municipality. Demographically these hotspot areas are characterised by low income and high unemployment.



**Figure 3:** (a) Distribution of the Robbery Index, (b) Distribution of the Contact Index, and (c) Distribution of the Murder Index

## 5. Conclusions

PCA is a very useful technique in deriving general and specific components of crime. Each of the different components of contact crimes exhibit unique patterns throughout the EMM. Violent crimes appear to be concentrated in the northern and southern parts of the study area, which are characterised by lower income, employment and education. Robbery typically occurs in the higher income areas of the EMM. This analysis can be expanded with the inclusion of the locations of all South African police stations in order to evaluate the spatial distributions of the components throughout the whole of SA. Further research opportunities exist for incorporating the underlying spatial autocorrelation in the PCA.

## References

- Ahaman, B. (1967). An analysis of crimes by the method of principal components. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **16**(1), 17-35.
- Burger, J., Gould, C. and Newham, G. (2010). The state of crime in South Africa: An analysis of SAPS crime statistics for 2009/10. *SA Crime Quarterly*, **34**, 3-12.
- Chainey, S., Tompson, L. and Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, **21**, 4-28.
- Holtmann, B. and Domingo-Swartz, C. (2008). Current trends and responses to crime in South Africa. Chapter 6: Crime, violence and injury prevention review. [http://www.mrc.ac.za/crime/cvi\\_second\\_review\\_ch6.pdf](http://www.mrc.ac.za/crime/cvi_second_review_ch6.pdf)
- Jolliffe, I.T. (2002). *Principal component analysis*. Second edition. Springer-Verlag, New York, Inc., 487 pages.
- Newham, G., Burger, J., Gould, C. And Lancaster, L. (2012). Fact Sheet: Explaining the official crime statistics for 2011/12. Institute for Security Studies. [http://www.issafrica.org/crimehub/siteimages/2012\\_Crime\\_Factsheet\\_20.09.12\[3.14 pm\].pdf](http://www.issafrica.org/crimehub/siteimages/2012_Crime_Factsheet_20.09.12[3.14_pm].pdf)
- Ratcliffe, J. (2010). *Crime mapping: spatial and temporal challenges*. In Handbook of Quantitative Criminology, Piquero, A.R. and Weisburd, D. (eds). Springer Science and Business Media, LLC, 2010.
- Salvati, L., Di Bartolomei, R., Rontos, K. and Bisi, S. (2012). Crime and the (Mediterranean) city: Exploring the geography of (in) security in Rome, Italy. *International Journal of Latest Trends in Finance and Economic Sciences*, **2**(1), 56-73.
- Stats SA. (2012). Victims of Crime Survey: 2012. Statistics South Africa. <http://www.statssa.gov.za/publications/P0341/P03412012.pdf>