International Journal of Criminology and Sociological Theory, Vol. 3, No. 2, June 2010, 477-495

Improved methods for surveying and monitoring crimes through likelihood based cluster analysis

Isak Neema[1]
Dankmar Böhning[2]

Abstract

This paper focuses on a development of a classification model that gives an accurate placement of regions into classes of the relative risk of crimes over time. The analysis was based on statistics on the cases of burglary and murder from 13 regions of Namibia for the period 2002 - 2006. Since crime statistics are counts, they are often contaminated by heterogeneity. The effect of population heterogeneity in the crime counts in particular makes comparison of crime risk across regions using traditional methods of classification impossible. As such a method for standardizing crime counts was introduced and models for modeling population heterogeneity proposed. In particular a mixture likelihood approach to clustering by McLachlan and Basford (1988) which was further extended for covariate effects was used. This is due to its ability in identifying important clusters and in mapping the relative risk of crime onto the study regions via the maximum a posteriori (MAP) method while inference was done via the EM algorithm of Dempster et al (1997). The result shows that the space - time mixture model conducted under non - parametric form gives a good account of the relative risk of the two crimes over time, while both space - time mixture and covariate adjusted space - time mixture models points to a 3 risk classification of the regional relative risk of the two crimes namely high, medium and low risk class respectively.

Introduction

The issue of increasing crime rate is a phenomenon affecting many countries in the world. It is due to increasing crime rate that scientific communities including statisticians are working together with law enforcement agencies in finding solutions to this issue. With much improved methods and techniques of crime surveillance in place, much of research work in

[1] University of Namibia, Department of Statistics, icekhuse@yahoo.com

[2] Dankmar Böhning, University of Reading, UK, d.a.w.bohning@reading.ac.uk

the field of Criminology has shifted towards analytic methods for analyzing crime surveillance data.

The issue of crime surveillance cannot be highly emphasized enough, as in the criminological studies surveillance is aimed at recording incidences of crimes as well as in early detection of possible changes in the patterns of crime distributions. In the case of Namibia in particular, crime surveillance can be used as a tool for effective policing and for initiating proper and timely crime intervention initiatives. The use of cluster analysis techniques in the classification of geographical areas or regions of interest is very useful in identifying areas of increasing crime rate. In particular, traditional methods of classification like single, average and complete linkage (Tsiamtsiouri and Panaretos; 1999) have been widely used. These methods use a Euclidean distance as a classification measure.

Although traditional methods have emerged as an important classification tools, they are not conducive in classifying areas from a heterogeneous setting. This is due to their sensitivity in dealing with crime data from areas of varying population sizes and lack of adjustment of the classification model for possible covariate effects. In addition, spatial clustering is not possible since the distance measure is only defined in a linear space while also highly skewed towards areas of high population sizes.

The crime data used in this paper comes from 13 regions of various population sizes for the period 2002 to 2006 (see figure 1). However, for comparison purposes, these data were complemented with statistics on the projected regional populations (population at risk) from the Central Bureau of Statistics (CBS). Therefore identification of regions with unusual high level of crime risk is necessary for initiating timely interventions, proper planning mechanism and better resource allocations.

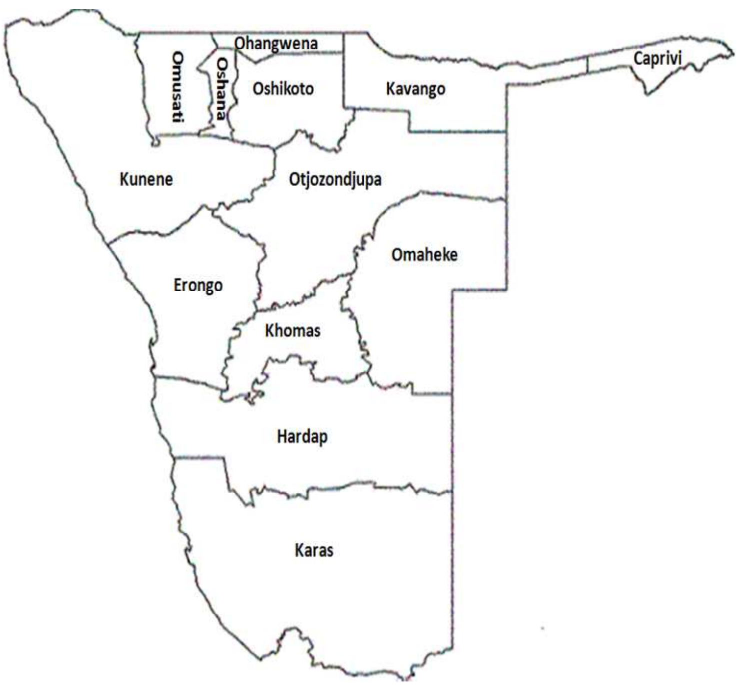


Figure 1: Regional boundary map of Namibia

Modeling Crime Counts

In a Criminological study, crime occurrence is measured in counts. As such, the Poison provides a natural distribution for modeling crime data. Although count data are associated with crime data, they often do exhibit over-dispersion (extra - Poison variation) violating the Poison assumption of equal mean-variance components. A diagnostic test for assessing over-dispersion or lack of thereof in the count data can be found in Böhning (1999). Therefore, alternative models dealing with over-dispersed count data have been proposed.

For instance Osgood (2000) propose using a negative binomial (NB) regression model in analyzing over-dispersed crime count data. Although this model is now widely used in the Criminological field, Berk and MacDonald (2008) argue that it does not always solve over-dispersion issue if fundamental errors in the regression model are not fully addressed. Alternatively, Böhning (2000 and 2003), McLachlan and Basford (1988) and Lawson et al (1999) introduce mixture models as an alternative to the NB.

Mixture models allow for unobserved complex structure in the data to be represented by a set of unobserved classes through probabilistic means (see also Böhning, Dietz and Schlattmann; 1998 and Lindsay; 1995). This leads to the application of mixture models in the area of cluster analysis (Aerts et al; 2002, McLachlan and Basford; 1988, and Everit; 1993), where it is used in the identification and allocation of areas into classes of homogeneous traits. According to Magidson and Vermunt (2002), recent developments in model based clustering provides improvements in the ability to identify these very important clusters and allocate areas into the classes accordingly.

Therefore, this paper focuses on a development of an appropriate classification model based on the concept of mixture models by McLachlan and Basford (1988) using an improved method of crime standardization and to propose suitable methods for monitoring changes in the regional crime risk over time.

Standardized crime ratio (SCR)

The presence of population heterogeneity in crime data necessitates the introduction of a standard method for standardizing the reported crime counts to the population at risk. This process enables smooth comparison of crime rates across study regions without significant loss of information. In particular, the SCR which is parallel to the standardize mortality ratio (SMR) implemented in Epidemiology and Public Health (see Böhning; 2004) is therefore introduced. Now, given the reported crime counts o_{rt} from region r during the time period t and the corresponding expected counts e_{rt}, we can define the SCR as

$$SCR_{rt} = \frac{o_{rt}}{e_{rt}}, \qquad r = 1,2,...,R \text{ and } t = 1,2,...,T \tag{1}$$

An interesting aspect from this equation is the calculation of the e_{rt}. Since the reference population is not readily available, the e_{rt} are calculated in terms of the population at risk n_{rt} as

$$e_{rt} = \hat{\lambda}_t n_{rt}$$

where, $\hat{\lambda}_t$ is the rate of crime during the time period t and it is defined as $\hat{\lambda}_t = \frac{\sum_r o_{rt}}{\sum_r n_{rt}}$. Fleiss et al (2003) refer to this technique as the internal indirect standardization method. One of the advantages of using this method of standardization according to Böhning (2004) lies in the marginal means of the ratio $\frac{\sum_r \sum_t o_{rt}}{\sum_r \sum_t e_{rt}}$ which is fixed to 1.

Finite mixture models

As before, given the observed crime counts o_{rt} and the expected counts e_{rt} calculated on the basis of internal indirect standardization, then information on the occurrences of crimes can be contained in the constructed SCR. Let λ_{jt} be the true parameter representing the relative risk of the SCR_{rt} in the j^{th} class during the time period t and assuming this relative risk is varying across study regions, then a population heterogeneity which is referring to the unobserved cluster variation in the parameter values of λ_{jt} exist.

Now, assuming that the o_{rt} follow a Poison distribution such that under SCR its distribution conditional on the model parameter is given by $o_{rt} \left| \lambda_t, e_{rt} \sim Po\left(o_{rt} \left| \lambda_t e_{rt}\right.\right)\right.$, then the conditional density of the o_{rt} takes the form

$$f\left(o_{rt} \left| P_t, e_{rt}\right.\right) = \sum_j Po\left(o_{rt} \left| \lambda_{jt} e_{rt}\right.\right) p_{jt}, \qquad j = 1, 2, ..., k \tag{2}$$

Equation (2) yield a k component finite spatial mixture of Poison distribution with a mixing density given by

$$P_t = \begin{pmatrix} \lambda_{1t} & \lambda_{2t} & \cdots & \lambda_{k_t t} \\ p_{1t} & p_{2t} & \cdots & p_{k_t t} \end{pmatrix} \tag{3}$$

implying that at the time period t, P_t is assigning proportion p_{jt} to cluster j with a relative risk λ_{jt}. In addition, p_{jt} should satisfy (i) $p_{jt} \geq 0$ and (ii) $\sum_j p_{jt} = 1$ for all t.

Although the model in (2) is good in presenting the spatial distribution of regional crime data over study areas, it is not ideal for monitoring crimes over time since comparison of the relative risk is not possible. This is because the model produces t independent mixtures which in time are incomparable as the number of components (k) of the mixture model might differ or if they are similar, the levels of spatial relative risk (λ_{jt}) might be substantially different. Hence regional allocation to the levels of the relative risk will not be entirely identical.

To overcome this problem, a mixture model that account for space and time components of the crime data by assuming fixed time periods is considered. In such a case, crime surveillance data from all time periods are viewed as a single data set. As such a single finite space - time mixture for the space - time crime data can be developed. Now, letting λ_j to be the unobserved cluster variation in the population such that $o_{rt} \mid \lambda_j, e_{rt} \sim Po\left(o_{rt} \mid \lambda_j e_{rt}\right)$, then we can re-write equation (2) as

$$f\left(o_{rt} \left| P, e_{rt}\right.\right) = \sum_j Po\left(o_{rt} \left| \lambda_j e_{rt}\right.\right) p_j, \qquad j = 1, 2, ..., k \tag{4}$$

where now $P = \begin{pmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_k \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}$. It is however observed from the above equation that λ_j does not depend on the time period t which is also reflected in the number of classes (k), while p_j should satisfy similar condition as p_{jt}. Furthermore, it is observed from the form of the mixing distribution that neither P_t nor P assumes any specific function which depends on the parameters to be estimated. Thus it is said to be in a non - parametric form.

Inference based on mixture models

The parameters of the mixing distribution are conventionally estimated using the non - parametric maximum likelihood estimators (NPMLE). The specific form of the log - likelihood for the space - time mixture is

$$l(P) = \prod_r \prod_t f\left(o_{rt} \mid P, e_{rt}\right) = \prod_r \prod_t \left\{ \sum_j Po\left(o_{rt} \mid \lambda_j e_{rt}\right) p_j \right\} \tag{5}$$

which is performed once as compared to the spatial mixture where the maximum likelihood is performed t times e.g.

$$\prod_t l(P_t) = \prod_r \prod_t \left\{ \sum_j Po\left(o_{rt} \mid \lambda_{jt} e_{rt}\right) p_{jt} \right\} \tag{6}$$

However, according to Böhning (2000) a closed form solution for the log - likelihood does not exist. As such the EM algorithm of Dempster et al (see Böhning; 1999 and 2000, Militino, Ugarte and Dean; 2001 and also the derivation of the EM algorithm in Chadrasekaran and Arivarignam; 2006) has been implemented to estimate the parameters of P when the number of components or classes k is fixed (known), else the vertex exchange method (VEM) (Lesperance and Kalbfleisch; 1992) that computes a grid of positive support points should precede the EM algorithm. The estimated parameters are given by $\hat{P} = \begin{pmatrix} \hat{\lambda}_1 & \hat{\lambda}_2 & \cdots & \hat{\lambda}_k \\ \hat{p}_1 & \hat{p}_2 & \cdots & \hat{p}_k \end{pmatrix}$.

In addition since the number of components k is not known, model selection criterions conducted within the maximum likelihood framework will be used in determining the best model fit as a number of possible candidate models will be fitted. Specifically, the Bayesian Information Criteria (BIC) by Schwartz (1988) and Akaike Information Criteria (AIC) by Akaike (1973) were used. For a k component space - time mixture model, the AIC and BIC value are calculated as follows

$$\begin{aligned} AIC &= -2\ln(l_k) + 2(2k-1) \\ BIC &= -2\ln(l_k) + (2k-1)\ln(R \times T) \end{aligned} \tag{7}$$

In the equations, $\ln(l_k)$ is the maximized log - likelihood value for the k component space - time mixture, $(2k-1)$ is the number of parameters to be estimated in the mixture model, while $(R \times T)$ is the number of data points. The model that minimizes the above criterions gives the best fit.

Crime mapping

A common practice in disease mapping and risk assessments is using a finite estimate of the non - parametric prior as a mapping tool (Böhning; 2000). Specifically, maximum a posteriori (MAP) estimation method by McLachlan and Krishnan (1997) and Rouse (2005) is used. Since for a given time period t the composition of $\hat{P}$ is formed by a set of disjoint clusters, then each region will only be assigned to one component of the relative risk. Therefore, for the space - time mixture in equation (4), the posterior probability for assigning regions to cluster j with a relative risk λ_j is

$$p\left(\lambda_j \mid o_{rt}, e_{rt}, \hat{P}\right) = \frac{Po\left(o_{rt} \mid \hat{\lambda}_j e_{rt}\right)\hat{p}_j}{\sum_l Po\left(o_{rt} \mid \hat{\lambda}_l e_{rt}\right)\hat{p}_l}, \qquad l, j = 1, 2, \ldots, k \tag{8}$$

Therefore in terms of MAP, a region will only be allocated to that cluster for which it has the highest posterior probability of belonging.

Application

The space - time mixture model discussed in the preceding sections was applied to the crime surveillance data from 13 administrative regions of Namibia for the 5 year time period (2002

- 2006). Specifically, space - time crime data on the cases of burglary and murder were considered. Table 1 gives the results of the fitted space - time mixtures and model selection criterions for the two crimes. In the table, k represent the number of components in the space - time mixture while $\log(l)$ gives the natural log - likelihood of the k component space - time mixture.

Table 1: Results of the fitted mixture and the model selection criteria

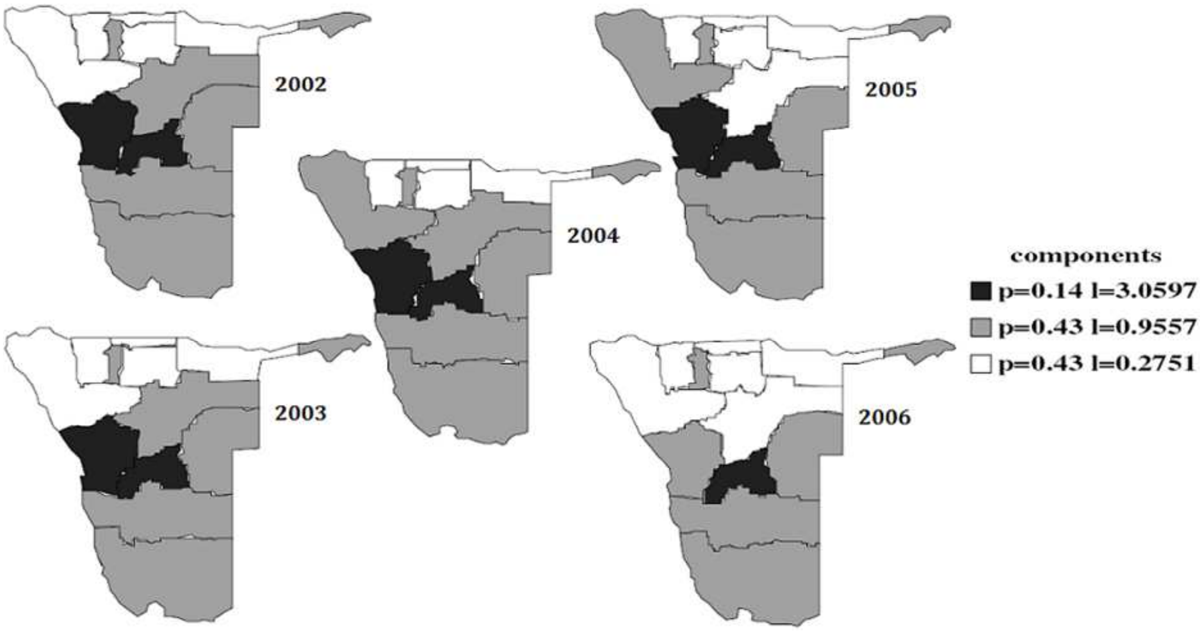
	Burglary			Murder		
k	$\log(l)$	BIC	AIC	$\log(l)$	BIC	AIC
4*	-169.016	367.253	346.032	-245.344	519.909	498.688
3	-169.016	358.904	344.032	-246.221	513.313	498.441
2	-175.635	363.793	355.270	-261.252	535.028	526.505
1	-187.919	380.012	377.838	-320.409	644.992	642.818

The result shows that for each crime, a 3 component space - time mixture give the best model fit as it minimizes both criterions (see also Appendix A.1 for the parameter estimates for each k component mixture). The parameter estimates for the crime specific mixing distributions for the 3 component space - time mixture are $\hat{P} = \begin{pmatrix} 0.275 & 0.956 & 3.060 \\ 0.43 & 0.43 & 0.14 \end{pmatrix}$ for burglary and $\hat{P} = \begin{pmatrix} 0.504 & 1.065 & 1.767 \\ 0.37 & 0.34 & 0.29 \end{pmatrix}$ for murder. These relative risk were compared to the threshold given by MLE for the relative risk in the homogeneous case $(\hat{\lambda} = 1)$.

This shows that only 14% of the space - time regions was assigned to the class with an increase in the relative risk of burglary of 3.060, while 43% apiece of the space - time regions was allocated to the class with a decrease in the relative risk of 0.956 and 0.275 respectively. Similarly in the case of murder, 37% of the space - time regions was allocated to the class with a decrease in the relative risk of 0.504 as oppose to 34% and 29% of the space - time regions that were allocated to the class with an increase in the relative risk of 1.065 and 1.767.

Practical interpretation of these statistics is that the space - time regional relative risk can be classified into three classes namely low, medium and high risk classes. Furthermore, regions in the high risk class are three times more at risk of burglary and nearly twice at risk of murder over time than they would normally be under homogeneous risk structures. The distribution of the space - time regional allocation to the components of the relative risk over time is presented in figure 2. From the figure, each map represent a time period, while legends signify the level of the relative risk which are low risk class (white color) to high risk class (dark grey) respectively.

(i) Burglary



(ii) Murder

Figure 2: Relative risk maps for the two crimes (2002 – 2006)

The maps show evidence of space - time clustering in the regional classification of the relative risk for the two crimes with mainly central and southern regions showing excess in the relative risk. In particular, Erongo and Khomas regions were consistently classified as high risk areas for burglary except in 2006 when Khomas was the only high risk area. Similarly in terms of murder, Karas and Khomas regions were classified as high risk areas over time, while Erongo was a high risk area only in 2003 and 2004. On the other hand, Omaheke region although consistently classified as a high risk area for murder, it has shifted to the medium risk class since 2005 while Hardap was the only region with fluctuating shift in all three components of the relative risk over time.

It is therefore understandably convenient to assume a north - south gradient in the case of murder while, in the case of burglary it is more difficult to determine where the direction of the greatest rate of increase in the relative risk lies, although a north - south - west gradient seems fairly justifiable.

Extended space - time mixture model

Here the space - time mixture model is further extended for possible covariate effects. In particular, the population density calculated as the ratio of regional population at risk by the size of the region was considered as a potential covariate. This is due to the fact that the Namibian Population and Housing census of 2001 indicated a high level of migration particularly from the northern to southern regions. In the case of mixture models, covariates are modeled through the mixed poison regression which according to McLachlan (1997) and Nylund and co (2007) gives a natural extension to the Poison regression model. This methodology is therefore extended to the space - time mixture Poison regression where covariates are included in the parameters of the mixing distribution. The mixed Poison regression is of the form

$$\begin{aligned}\log\left(\lambda_{rtj}\right) &= \eta_{rtj} \\ &= \log\left(e_{rt}\right)+\beta_{0j}+\beta_1 x_{rt}, \qquad j=1,2,\dots,k \end{aligned} \tag{9}$$

This is a general form of a linear mixed model (GLMM) with a k component space - time mixture over the intercept β_0 and a fixed effect over the covariate variable. Furthermore, equation (9) is also referred to as the partial random effect model (PREM). Although the regression model in equation (9) has different intercepts for each j component, the covariate effect remains identical for each component. Similarly, the k component space - time mixture can also be allocated to the covariate effect and then evaluate the model fit. This gives a full random effect model (FREM) which is of the form

$$\eta_{rtj} = \log\left(e_{rt}\right)+\beta_{0j}+\beta_{1j} x_{rt}$$

Parameter estimates for the space - time mixture Poison regression is done via adaptation of the EM algorithm. In this case, the marginal distribution of the o_{rt} is

$f\left(o_{rt} \mid P,\eta_{rtj}\right)=\sum_j Po\left(o_{rt} \mid e^{\eta_{rtj}}\right)p_j$ while the corresponding log - likelihood function is now of the form $l(P)=\prod_r\prod_t\sum_j Po\left(o_{rt} \mid e^{\eta_{rtj}}\right)p_j$. The parameters of the mixing distribution are therefore estimated to be $\hat{P}=\begin{pmatrix}\hat{\psi}_1 & \hat{\psi}_2 & \cdots & \hat{\psi}_k \\ \hat{p}_1 & \hat{p}_2 & \cdots & \hat{p}_k\end{pmatrix}$, where $\hat{\psi}_j=\left(\hat{\beta}_{0j},\hat{\beta}_{1j}\ or\ \hat{\beta}_1\right)$. This implies that $\hat{P}$ is assigning weights $\hat{p}_j$ to the regression parameters $\hat{\beta}_{0j}$ and $\hat{\beta}_{1j}$ in the case of FREM or $\hat{\beta}_{0j}$ and $\hat{\beta}_1$ in the case of PREM. Since the number of components k of the space - time mixture regression is not known, a set of candidate models will be fitted. Once again the AIC and BIC criterions will be used in determining the best model fit. These criterions are now defined as $AIC=-2\ln(l)+4k$ and $BIC=-2\ln(l)+(2k)\ln(R\times T)$ under PREM or $AIC=-2\ln(l)+2(3k-1)$ and $BIC=-2\ln(l)+(3k-1)\ln(R\times T)$ in the case of FREM. Similarly, the posterior probability of assigning regions to the components of the space - time mixture regression is

$$\hat{p}_{rtj}=\frac{Po\left(o_{rt} \mid e^{\eta_{rtj}}\right)\hat{p}_j}{\sum_l Po\left(o_{rt} \mid e^{\eta_{rtl}}\right)\hat{p}_l}, \qquad l,j=1,2,\ldots,k$$

(10)

The outcome of the model fit is shown in table 2. For each crime, the result shows no significant improvement in the log - likelihood of the PREM with a 3 and 4 component space - time mixture over the intercept. As a result a FREM with a 3 component space - time mixture over both intercept and population density effect was fitted. This undertaking lead to a considerable improvement in the log - likelihood in the case of burglary only.

Table 2: Results of the fitted mixtures and the model selection criteria

	Burglary			Murder		
PREM k	$\log(l)$	BIC	AIC	$\log(l)$	BIC	AIC
1	-198.045	404.439	398.020	-329.357	667.063	660.714
2	-167.501	351.700	339.002	-252.761	522.220	509.522
3	-161.322	347.690	328.644	-243.595	512.236	493.190
	-161.322	356.039	330.644	-242.764	518.923	493.528

4						
FREM $k=3$	-157.182	347.759	320.364	-243.108	519.611	492.216

Furthermore, the model selection criterion indicates that a 3 component model is supported in the case of burglary. This result is also confirmed by the estimated parameters of the covariate adjusted space - time mixture model in Appendix A.2. In this case the parameter estimates for a 4 component model shows two components with nearly identical relative risk of 1.314 and 1.311. Although this values seems different under more decimal places, they are treated equal and hence were merged and given an average relative risk of 1.313. This results in a model with a 3 component space - time mixture over the intercept, a result which coincides with the model that minimizes the two criterions.

The preferred model in the case of burglary shows no significant difference between the model under PREM (BIC = 347.690) to that under FREM (BIC = 347.759), although with respect to the AIC a model with a 3 component space - time mixture under PREM is somehow weekly supported. Hence a 3 component model under PREM was used for the rest of the analysis as it minimizes both criterions and we have less parameter to estimate. It can also be shown that the effect of population density on the risk of burglary will somehow be very close if not identical in the case of FREM, hence it is irrelevant as to which of the two models is chosen. Similar conclusion can also be made in the case of murder, where a 3 component space - time mixture under PREM minimizes the two criterions.

Residual heterogeneity is also detected when the space - time mixture is adjusted for population density as the resulting log - likelihood of the preferred model under the two criterions improves by 7.694 (burglary) and 2.626 (murder) as compared to the log - likelihood of the corresponding 3 component space - time mixture in table 1. The estimated parameters of the mixing distribution of the space - time mixture regression and the solution under homogeneous case $(k=1)$ is given in table 3.

Table 3: Parameter estimates of the 3 component covariate adjusted space – time mixture under the partial random effect model (PREM)

Crime	k	$\hat{p}_j$	$\hat{\lambda}_j$	$\hat{\beta}_{0j}\,(S.E.)$
	1	1.00	1.030	0.190 (0.075)
Burglary	3	0.15	3.854	1.938 (0.048)
		0.13	1.309	0.858 (0.071)
		0.72	0.387	-0.361 (0.122)
	1	1.00	1.031	0.161 (0.035)

Murder	3	0.15	1.877	0.709 (0.032)
		0.40	1.282	0.327 (0.036)
		0.45	0.567	-0.489 (0.050)

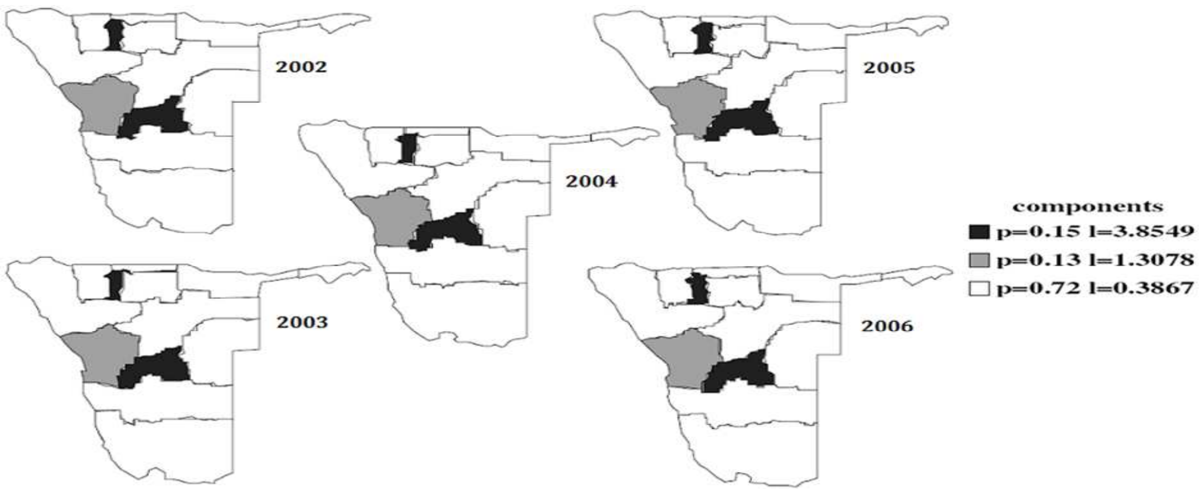
The proportion of regional space - time data assigned to the high risk classes are of interest. The result from the table shows that 15% of the regional space - time data was allocated to the high risk class in both cases with an intercept parameter of 1.938 in the case of burglary and 0.709 in the case of murder. A further detail comparison of the relative risk between the space - time mixture model (STM) and the covariate adjusted space - time mixture (CASTM) is presented in table 4. In the table, $\Delta\hat{p}_j$ is the change in the proportion of regional space - time data in the j^{th} class, while $e^{(\hat{\beta}_{0j})}$ and $e^{(\hat{\beta}_1)}$ are the changes in the relative risk of the two crimes in the j^{th} class when population density is negated and the effect of the population density itself.

Table 4: Comparison of the risk classes between the space – time mixture model and the population density adjusted space – time mixture model.

Crime	**Risk class**	$\hat{p}_j$		$\Delta\hat{p}_j$	$e^{(\hat{\beta}_{0j})}$	$e^{(\hat{\beta}_1)}$
		STM	**CASTM**			
Burglary	$k=1$	-	-	-	1.209	0.973
	High	0.14	0.15	0.01	6.945	
	Medium	0.43	0.13	-0.30	2.358	0.907
	Low	0.43	0.72	0.29	0.697	
Murder	$k=1$	-	-	-	1.175	0.978
	High	0.29	0.15	-0.14	2.032	
	Medium	0.34	0.40	0.06	1.387	0.987
	Low	0.37	0.45	0.08	0.613	

The result shows that after adjusting for population density the proportion of regional space - time data in the high risk class increases by 1% with those in the low risk class increasing by as much as 29% in the case of burglary. This points to a high level trade - off in the proportion of regions in the medium class with a high percentage shifted to the low risk class. However in the case of murder, a 14% decrease in the proportion of regional space - time data in the high risk class is observed while in the low risk class it increases by a mere 8% .It is further observed from table 4 that regions in the high risk class are six times more at risk of burglary than they would normally be under homogeneous case when population density is negated. This is double the risk in the case of the space - time mixture model. Similarly in the case of murder, when the effect of population density is negated regions in the high risk class are twice at risk for murder which is somehow similar to the risk in the space - time mixture model. The distribution of the space - time regions into components of the above relative risk is presented in figure 3 below.

(i) Burglary



(ii) Murder

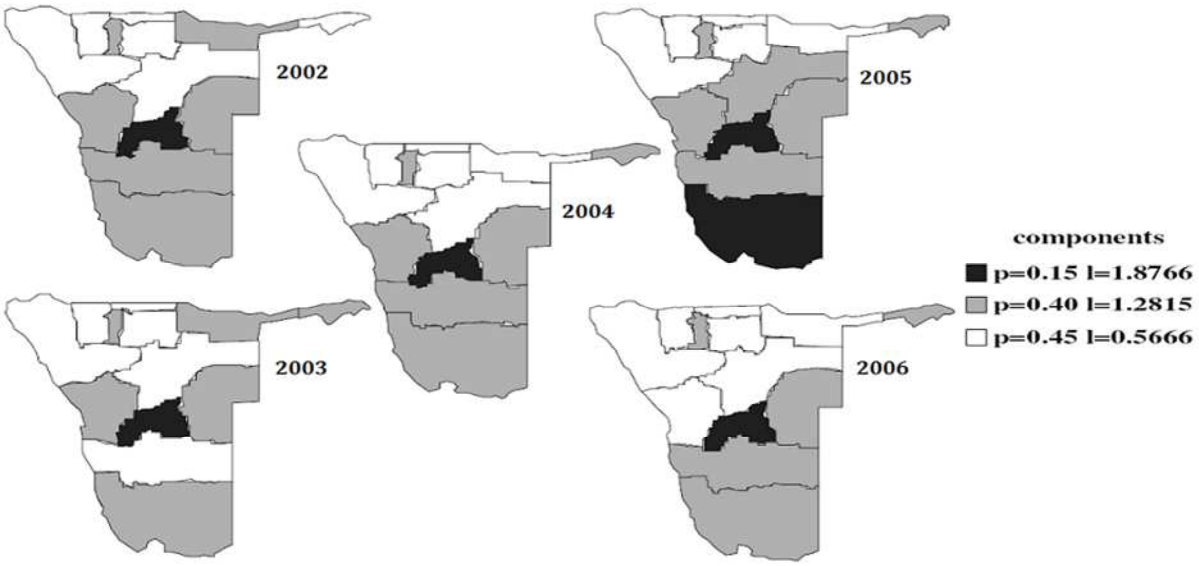


Figure 3: Relative risk maps for the two crimes (2002 - 2006) accounting for population density

Comparison of the relative risk maps in figure 2 and 3 shows a clear shift in the relative risk of some regions in both cases. Particularly in the case of burglary, Erongo region shifted from a high to the medium risk area while Oshana shifted from a medium to the high risk area when the population density is taken into account. Similarly in the case of murder, Karas region shifted from a high to a medium risk area although it was classified as a high risk area in 2005. In addition, Erongo region which has been classified as a high risk area in 2003 and 2004 and a medium risk area in 2006 has now shifted to the medium and low risk classes respectively. Furthermore, Hardap which was a high risk area in both 2002 and 2006 periods is now classified as a medium risk area in both cases when the population density is taken into account, while Omaheke region has now largely become a medium risk area for murder throughout the study period.

It is therefore of interest to note that the differences in the allocation of regions in the relative risk maps of figure 3 can be attributed to the high value of the intercept term in the regression model. This is because the effect of population density is quite minimal across the relative risk classes in both cases. It is also observed from table 4, that the effect of population density for each crime is similar to the effect in the respective homogeneous models. This is also a further indication that the population density is somehow not significant in the model.

Overall, the maps in figure 2 and 3 clearly show evidence of spatial dependency in the regional relative risk for each crime. Although the magnitude of the degree of association is not clear at this stage, it can be deduced from the two figures that geographical characteristics of the regions seems to play an important role as most of the regions are consistently allocated to the same components of the relative risk over time.

Discussion

This paper has focused on a development of a classification model that provides a more accurate classification of regions into classes of the relative risk of crime based on the specific regional relative risk over time. The analysis was based on secondary data set of the statistics on the cases of burglary and murder from the 13 regions of Namibian complemented with statistics on the regional population projections from the National planning commission.

One of the important contributions in this paper is the introduction of the standardized crime ratio (SCR) in section 3, from which the proposed classification model and maps of the regional relative risk are based. This ratio enable direct comparison of the relative risk of crimes from regions with varying population sizes an effect refers to as the population heterogeneity in the paper. In the field of Criminology, this effect has been largely overlooked in the analysis of crime and risk assessments, which usually leads to high inflation or overestimation in the point estimates of the relative risk of crimes. Hence the SCR was introduced to absorb this effect by standardizing the crime statistics to the population at risk.

A mixture likelihood approach under model based clustering was proposed as a classification model. In particular the non - parametric form of the space - time mixture model was found to give a good account of the relative risk of the two crimes over time. This model was further extended to allow for the possible effect of population density on the regional classification referred to as the covariate adjusted space - time mixture model.

Both the space-time mixture and covariate adjusted space-time mixture model indicates a 3 risk classification of the space-time regional relative risk of the two crimes namely high, medium and low risk classes. Furthermore, the regional allocation to the components of the relative risk shows an increasing north-south gradient in the case of murder while a north-east-south gradient in the case of burglary is fairly justifiable. It is further observed that when the population density is negated, the risk of burglary in the regions allocated to the high risk class doubled as compared to that of the space-time mixture, while in the case of murder it largely remain constant between the two models. In addition, population density was not found to be significantly affecting the regional classification as the high risk classes in both cases constitute regions with varying population densities.

Appendix A

A.1: Parameter estimates for the space – time mixture model

Components k	**Burglary**		**Murder**	
	p_j	λ_j	p_j	λ_j
4	0.30	0.275	0.36	0.500
	0.13	0.275	0.29	1.008
	0.43	0.956	0.27	1.572
	0.14	3.060	0.08	2.010
Merging equal estimates	-	-		
	0.43	0.275		
	0.43	0.956		
	0.14	3.060		
3	0.43	0.275	0.37	0.504
	0.43	0.956	0.34	1.065
	0.14	3.060	0.29	1.767
2	0.80	0.464	0.47	0.560
	0.20	2.810	0.53	1.515

A.2: Parameter estimates for the covariate adjusted space – time mixture model

Components k	Burglary		Murder	
	p_j	λ_j	p_j	λ_j
	0.72	0.387	0.35	0.529
4	0.07	1.314	0.15	0.761
	0.06	1.311	0.39	1.369
	0.15	3.851	0.11	1.927
Merging equal estimates	-	-		
	0.72	0.387		
	0.13	1.313		
	0.15	3.851		
	0.72	0.387	0.45	0.567
3	0.13	1.309	0.40	1.282
	0.15	3.854	0.15	1.877
2	0.78	0.445	0.50	0.599
	0.22	3.003	0.50	1.571
FREM	0.57	0.359	0.45	0.566
$k=3$	0.24	0.795	0.36	1.249
	0.19	2.907	0.19	1.845

References

Aerts, M., Geys, H., Molenberghs, G. and Ryan, L.M. (2002). Topics in modeling of clustered data, 1^{st} edition. Boca Raton, Florida: Chapman and Hall.

Beck, R. and MacDonald, J. (2008). Overdispersion and Poisson regression. Journal of Quantitative Criminology, Vol. 24, No. 3, pp: 269 - 284, doi: 10.007/s10940 - 008 - 9048 - 4.

Böhning, D. (2003). Empirical bayes estimators and non - parametric mixture models for space and time - space disease mapping and surveillance. Environmetrics, 14: 431 - 451.

Böhning, D. (1999). Computer - assisted analysis of mixtures and applications: Meta-analysis, disease mapping and others. Boca Raton: Chapman and Hall.

Böhning, D. (2003). The EM algorithm with gradient function update for discrete mixtures with known (fixed) number of components. Statistics and Computing, 13, pp: 257 - 265.

Böhning, D., Dietz, E. and Schlattmann, P. (1998). Recent developments in computer-assisted analysis of mixtures. Biometrics. 54: 525 - 536.

Böhning, D., Dietz, E. and Schlattmann, P. (1998). Recent developments in computer assisted analysis of mixtures (C.A.MAN). Biometrics. 54: 367 - 377.

Böhning, D., Sarol, J., Rattanasiri, S., Viwatwongkasem, C. and Biggeri, A. (2004). A comparison of non - iterative and iterative estimators of heterogeneity variance for the standardized mortality ratio. Biostatistics. Vol. 5, No. 1, pp: 61 - 74.

Chandrasekaran, K. and Arivarignan, G. (2006). Disease mapping using mixture distribution. Indian Journal of medical Research, 123, pp 788 - 798.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B (Methodological), Vol. 39, No. 1, pp: 1 - 38.

Everit, B.S. (1993). Cluster Analysis, 3^{rd} edition. London: Edward Arnold.

Fleiss, J.L., Levin, M. and Paik, M.C. (2003). Statistical Methods for rates and proportions, 3^{rd} edition. Hoboken, New Jersey: John Wiley and Sons.

Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.F and Bertollini, R. (1999). Disease mapping and risk assessment for public health. New York: John Wiley and Sons.

Lesperance, M.L. and Kalbfleisch, J.D. (1992). An algorithm for computing the nonparametric MLE of a mixing distribution. Journal of the American Statistical Association, Vol. 87, No. 417.

Lindsay, B.G. (1995). Mixture models: Theory, Geometry and Applications. NSF - CBMS Regional Conference Series in Probability and Statistics, Volume 5.

Magidson, J. and Vermunt, J.K. (2002). Latent class models for clustering: A comparison with K - means. Canadian Journal of Marketing Research, Vol. 20.

McLachlan, G.J. and Basford, K.E. (1988). Mixture models: Inference and applications to clustering, Volume 84. New York: Marcel Dekker.

McLachlan, G.J. (1997). On the EM algorithm for overdispersed count data. Statistical Methods in Medical research, 6: 76 - 98.

McLachlan, G.J. and Krishnan, T. (1997). The EM algorithm and Extensions. New York: John Wiley and Sons.
Militimo, A.F., Ugarte, M.D. and Dean, C.B. (2001). The use of mixture models for identifying high risks in disease mapping. Statistics in Medicine. 20: 2035 - 2049. doi: 10.1002/sim.821.
Nylund, K.L., Asparonhov, T. and Muthen, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study. Structural Equation Modelling, Vol. 14, No. 4, pp: 535 - 569.
Osgood, D.W. (2000). Poisson - based regression analysis of aggregate crime rates. Journal of Quantitative Criminology; Vol. 16, No. 1.
Rouse, D.M. (2005). Estimation of finite mixture models. Master of Science thesis. North Carolina State University, North Carolina, USA.
Tsiamtsiouri, A. and Panaretos, J. (1999). Some statistical analysis of Greek crime data. Munich Personal RePEc Archive paper, No. 6353, posted 18.