Ecological Analysis of small area offence and offender data

Robert Haining

Department of Geography
University of Cambridge

Seminar: Security Matters! Stockholm, May 20-21

Outline:

1. The nature of ecological analysis.
2. Three examples:
    1. Analysing geographical variation in burglary rates.
    2. Area profiles and high intensity crime areas.
    3. Analysing the geographical distribution of offenders and estimating the risk of offending.
3. The challenges presented by ecological analysis.
4. Final points.

1. The nature of ecological analysis.

Ecological: study of groups or aggregates using data grouped by:

social class;
socio-economic status;
demographics (sex, age cohorts) ....

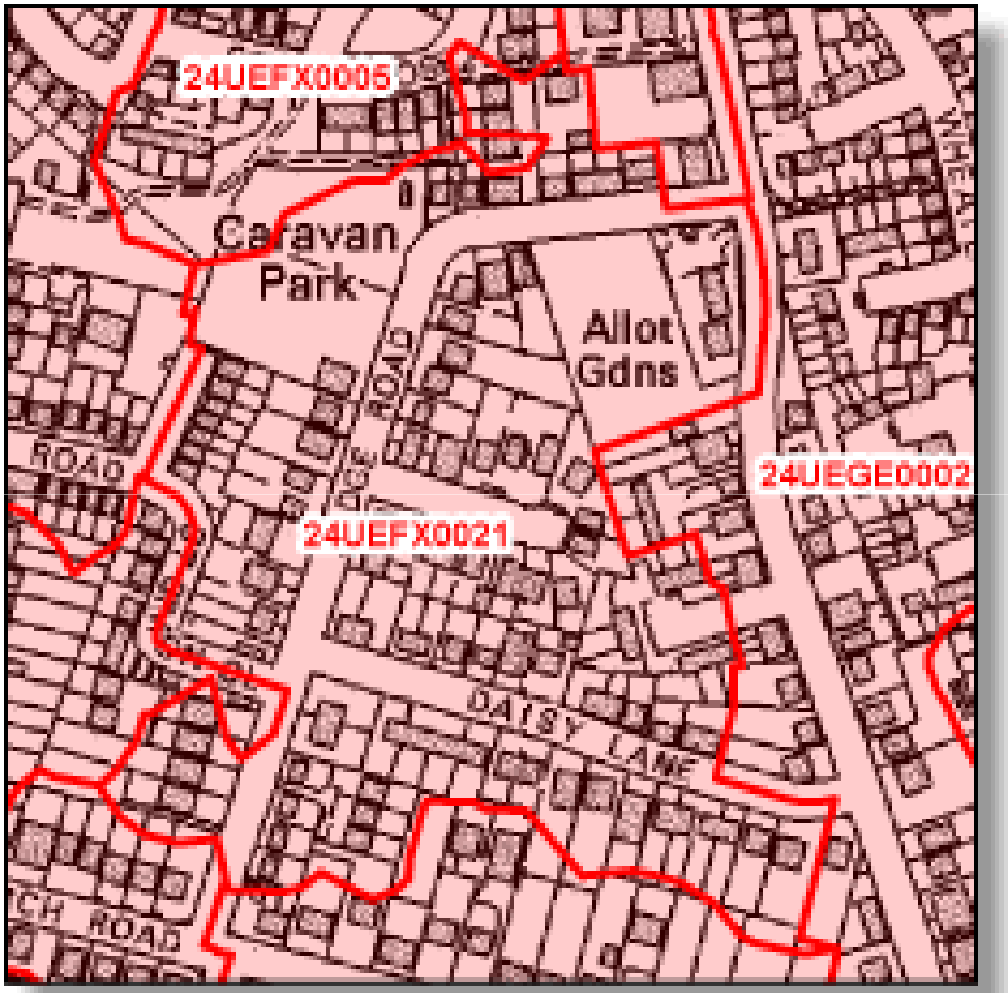
geography (and time).

Geography  scale

UK Census: Census Output Areas (COAs) – c. 220,000.

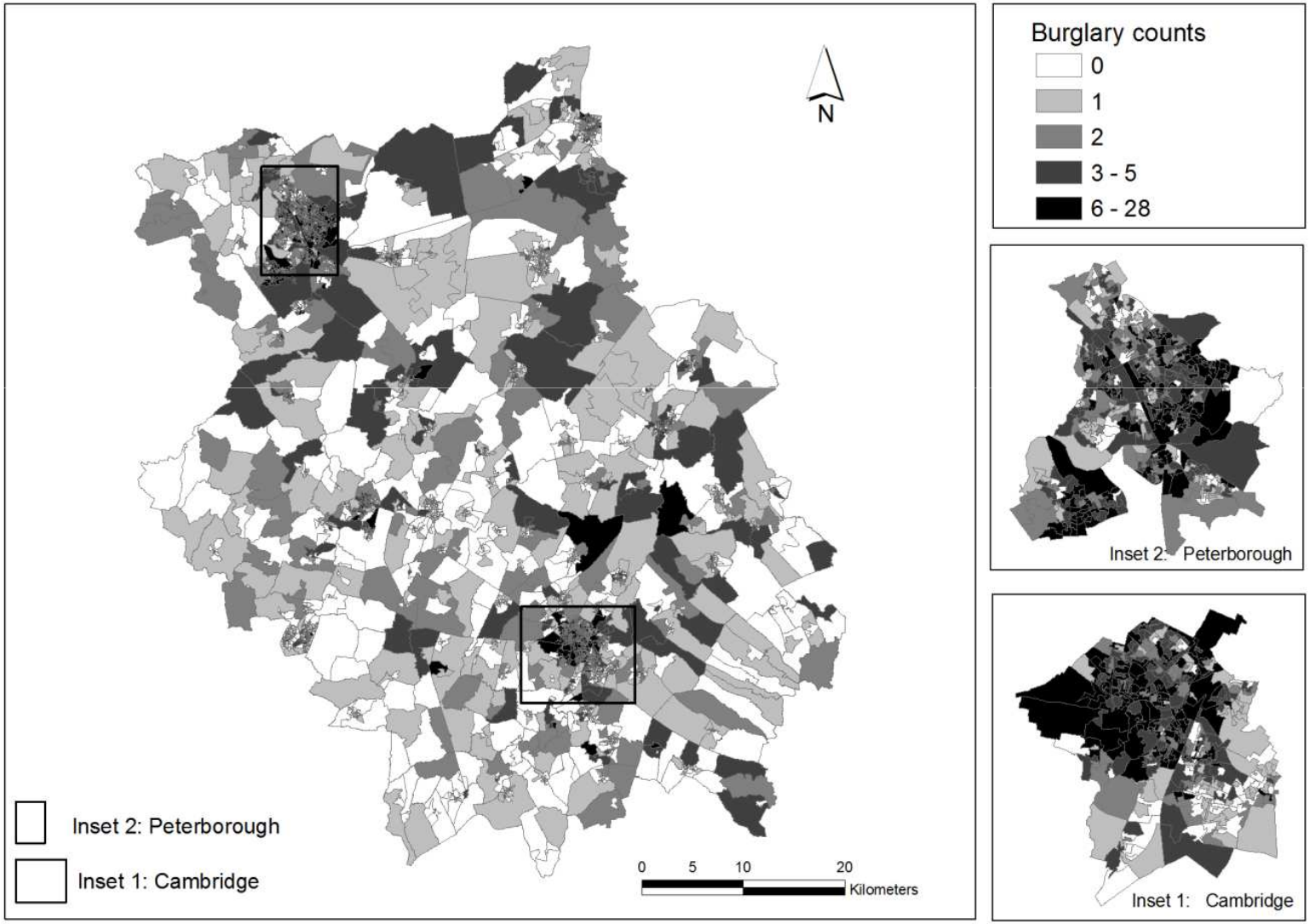
Design criteria for COAs:

- aggregations of postcodes;
- recommended to contain approx. 125 households;
- socially homogeneous based on housing tenure and dwelling type.


24UEFX0005
Caravan Park
Allot Gdns
24UEGE0002
24UEFX0021
DAISY LANE

(ONS website)

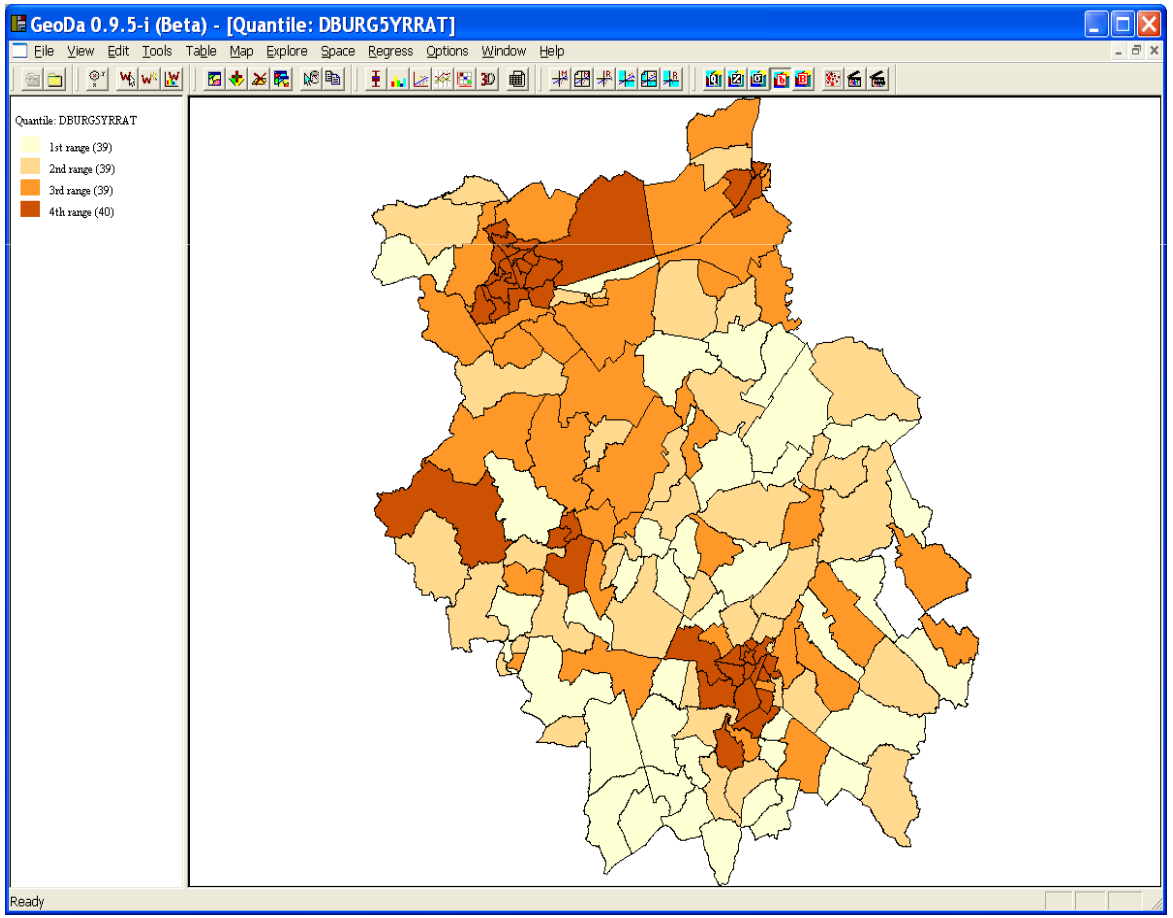
Quantile map of burglary counts at COA level in Cambridgeshire, 2002.



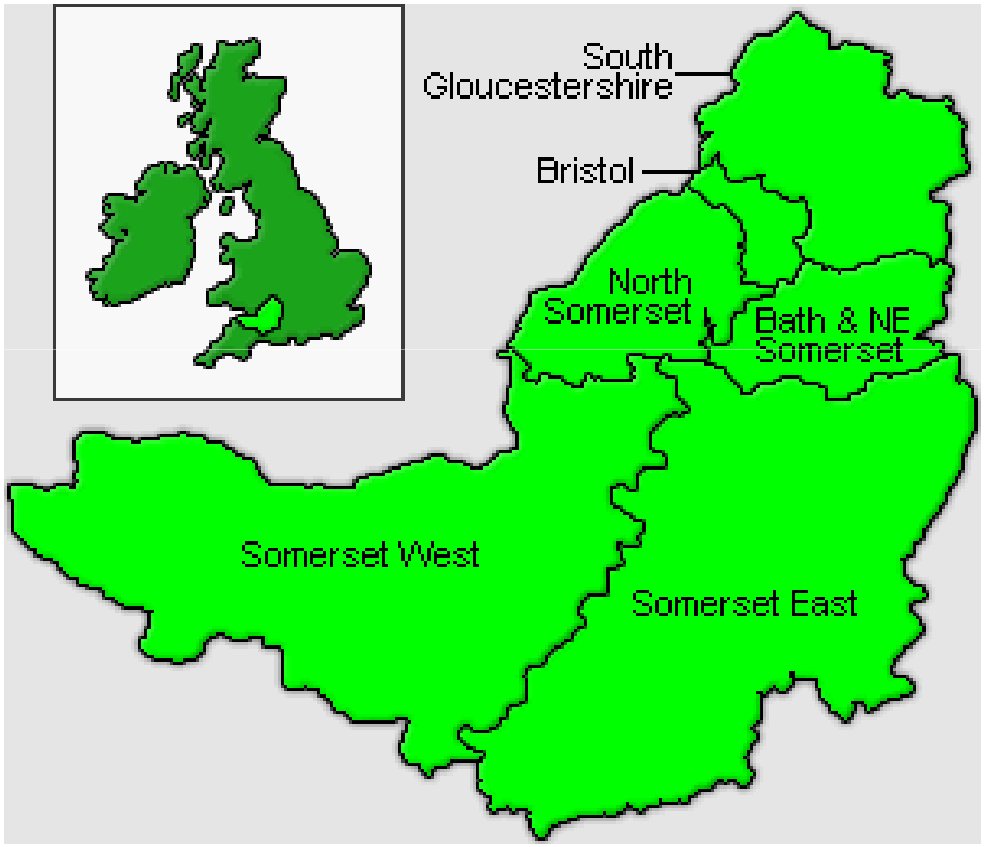
The next level up is the Super Output Area (SOA) with 3 levels:

- lower layer SOA (LSOA: minimum population size 1,000)
- middle layer SOA (MSOA: minimum population size 5000)
- upper layer SOA (USOA: minimum population size 25000)

Burglary rates for Cambridgeshire MSOAs (0.55m population; c.150 MSOAs)



In UK: 52 Police Force Areas (PFAs)



Avon and Somerset PFA

Recorded offence and offender data made available by postcode area and COA (offence location; victim address; offender address) .

POSTCODE_A	COA
CB58SX	12UGJD
CB19HU	12UBFT
PE13BL	00JANS
PE13BL	00JANS
PE13BL	00JANS
PE13UL	00JANM

Data quality issues:

[1] Accuracy (geocoding; time of event)

REPORTED	OCCURRED_FROM	OCCURRED_TO
05-Jul-05	01-Jul-05	04-Jul-05
05-Jul-05	04-Jul-05	05-Jul-05
05-Jul-05	05-Jul-05	05-Jul-05
05-Jul-05	04-Jul-05	05-Jul-05
05-Jul-05	02-Jul-05	03-Jul-05

[2] Completeness (e.g. offender not known; not all offences reported; limited personal details)

[3] Consistency.

[4] Resolution.

Forms of ecological analysis:

Descriptive: maps (presentation graphics; visualization tools); graphical and numerical summaries (e.g. hotspot locations).

Confirmatory: model fitting for parameter estimation and hypothesis testing.

Why are ecological analyses of crime data useful and important?

[a] Police are “territorial” and one aspect of resource allocation is by geographical area. PFAs, BCUs and beats/ neighbourhoods.

[b] Many theories about the location of offences and where offenders live have an ecological level:

- what is the appropriate spatial framework?

- what is relationship between the appropriate framework and data availability?

- role of ecological analysis when the target of inference is the individual rather than the area.

2. Some examples of ecological analysis:

2.1 Analysing geographical variation in burglary rates.

The ecological dimension: the act of burglary as the outcome of a rational choice “two stage process”.

Stage 1: Select area
Stage 2: Select target within the chosen area.

Each stage involves a distinct set of factors

Area selection (Bernasco and Luykx, 2003):

- *attractiveness* (reward): affluent areas favoured over less affluent and deprived areas.
- *opportunity* (risk): areas with fewer formal and/or informal capable guardians offer a greater likelihood of success.
- *accessibility* (familiarity + least effort principle): areas which are known to the offender but where (s)he will be anonymous and which are nearby.

For the motivated offender, choice of area, is a balance of risk against reward whilst taking into account the effort involved.

Questions:

[1] What is the significance of each of these three sets of factors and how far do they help us to explain area differences in burglary rates.
{Modelling for the purpose of hypothesis testing}

[2] By how much, on average, do area level rates of burglary increase for unit increases in each of the different factors.
{Modelling for the purpose of parameter estimation}

Bernasco and Luykx (2003): Study of burglary rates 1996-2001 for 89 residential neighbourhoods in The Hague, Netherlands.

Each neighbourhood contained c. 5000 people. 26,000 burglaries or on average 50 per neighbourhood.

Attractiveness variables: homeownership rates; property values.

Opportunity variables: ethnic heterogeneity; residential mobility rates.

Accessibility variables: accessibility index (based on residential addresses of convicted burglars); proximity to the CBD.

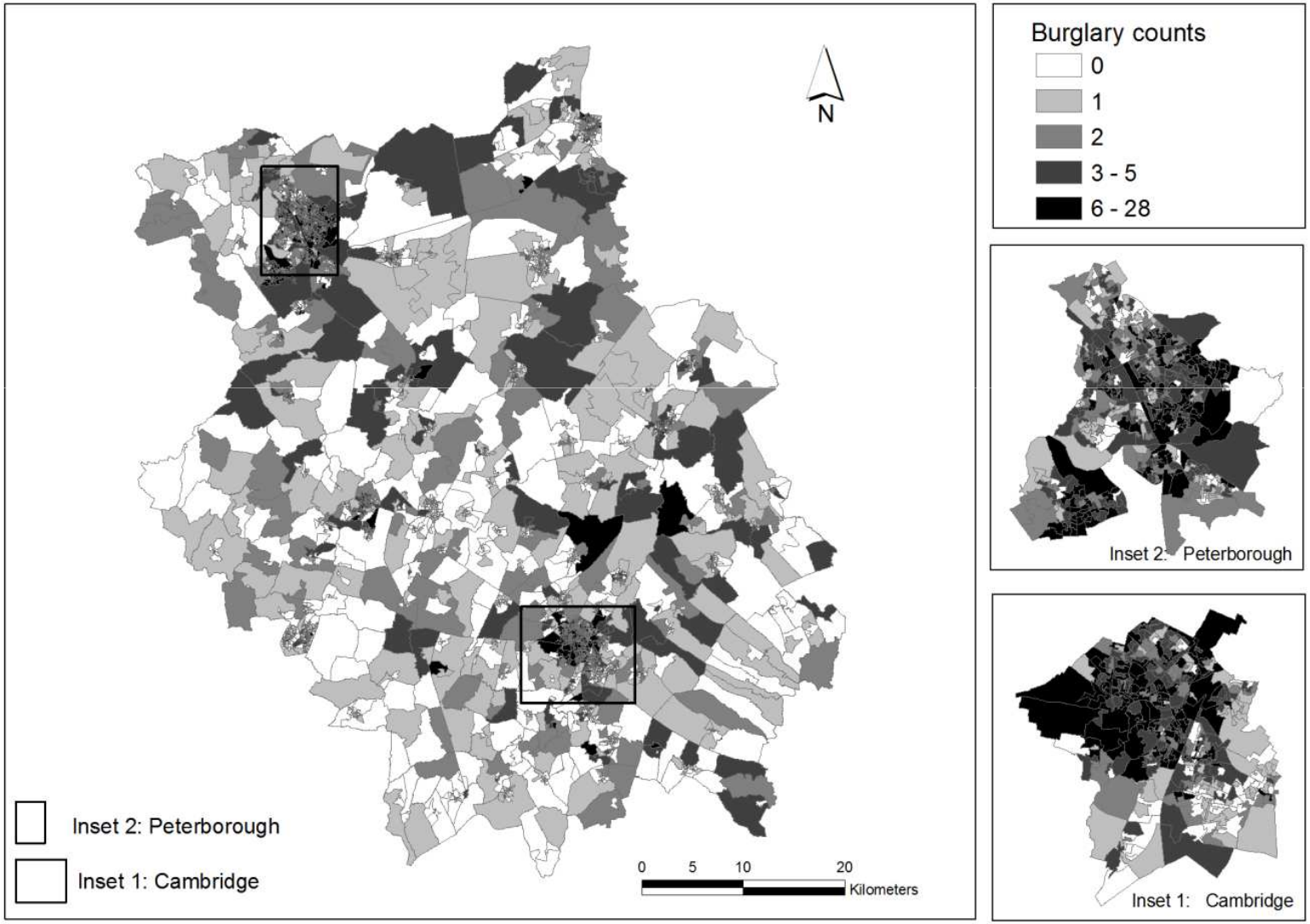
Multivariate regression analysis of burglary rates in The Hague neighbourhoods: standardized regression coefficients, t values and significance levels. (Bernasco and Luykx, 2003.)

Independent variable	**Standardized regression coefficient**	**t-value**
Accessibility index (Ac)	0.39**	6.03
Proximity to central business district (Ac)	0.14*	2.19
Residential mobility (Op)	0.30**	4.31
Ethnic heterogeneity (Op)	0.36**	4.07
Property values (Att)	0.24**	4.02
Home ownership (Att)	0.17*	2.61

* $p<0.05$ (2 tailed test).
** $p<0.01$ (2 tailed test).

R^2 (model fit)= 0.84

Quantile map of burglary counts at COA level in Cambridgeshire, 2002.



Multiplicative change in the expected number of burglaries, due to unit increase in the corresponding variable: Cambridgeshire COAs 2002.

	Negative binomial GLM
detach_h[a] [Att]	1.0026
rent_priva[a] [Att]	1.0112
lonep_h[a] [Op]	1.0286
eInactive[a]	1.0170
m_nstud_h[a]	1.0363
no_couple[a] [Op]	1.0116
Accessq3[b] [Ac]	1.4978
Accessq4[b] [Ac]	2.3560
Peterborough[c]	1.2827

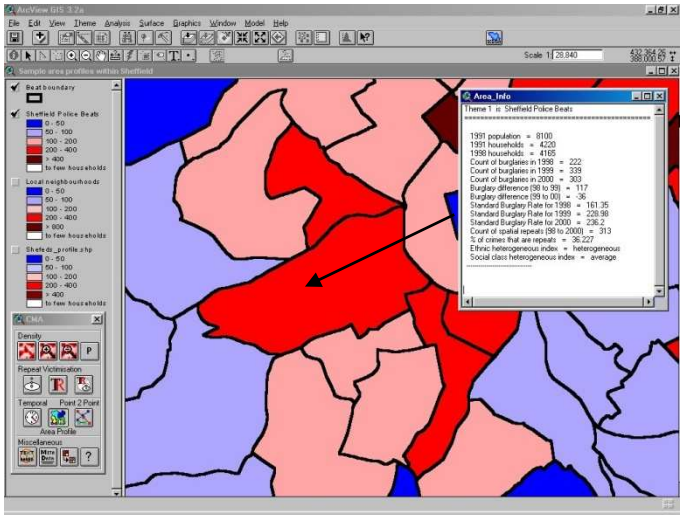
[a] increase in the expected count for a one percentage point increase in covariate.

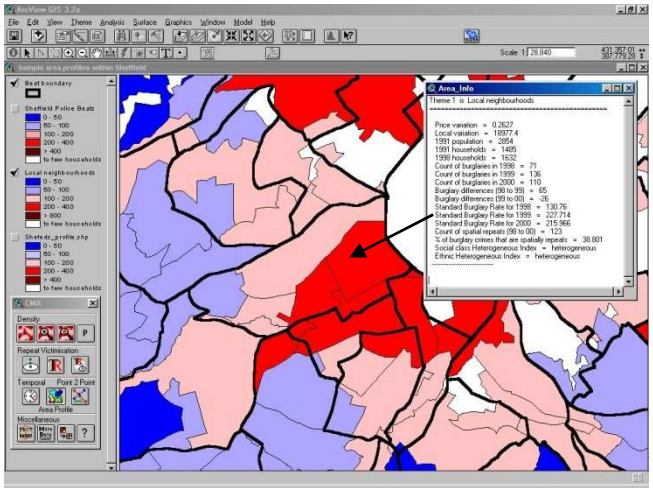
[b] increase in the expected count compared to areas below the median accessibility score.

[c] increase in the expected count for any COA in Peterborough compared to the rest of the county.

Haining and Law (2010)

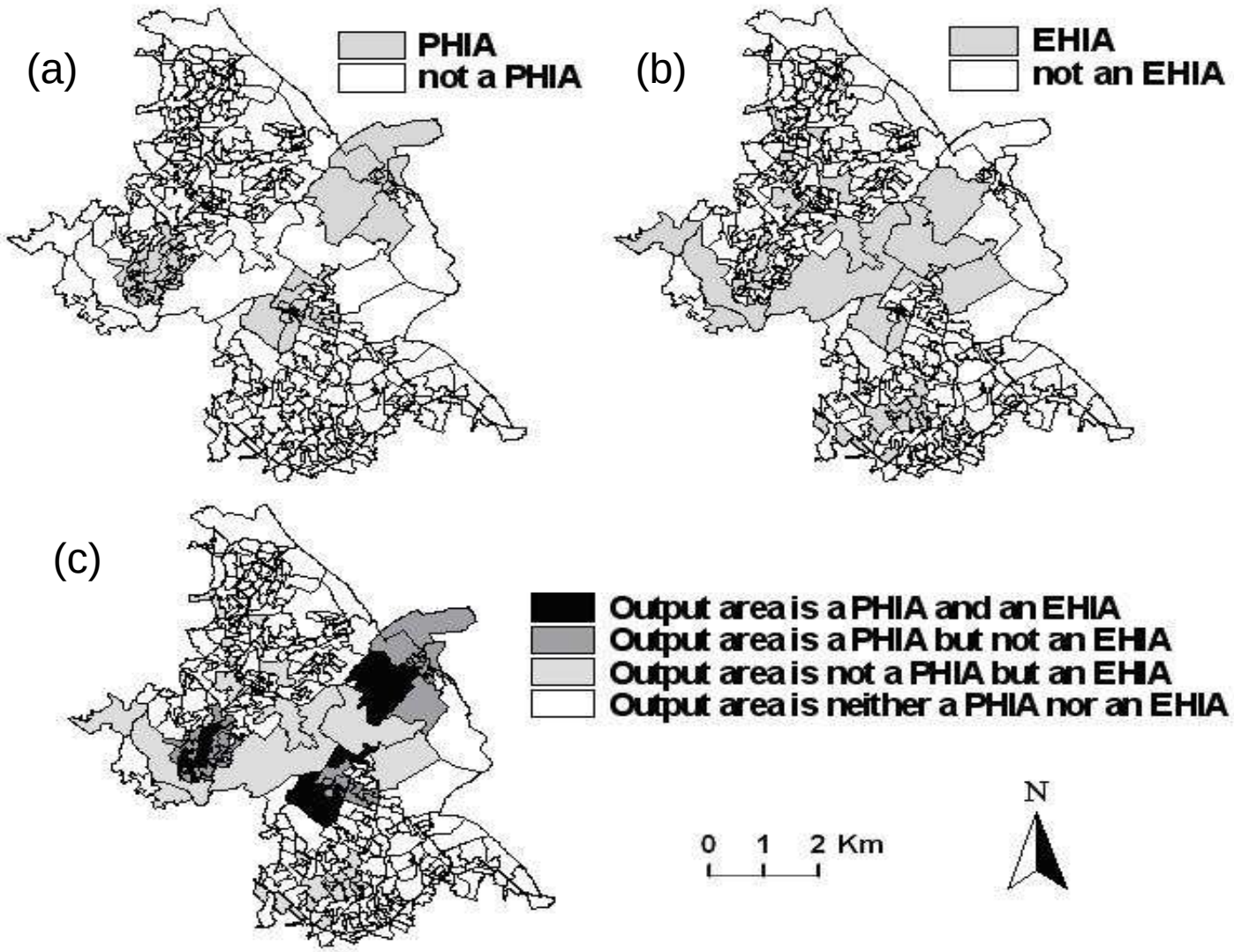
2.2 Area profiles and high intensity crime areas.





Different spatial frameworks (police beats and neighbourhood areas) used to compare Standardised Burglary Rates for an area in Sheffield in 2000.

Profile data on selected areas displayed using “point and click” operation in a GIS. (Brindley et al. 2008)



Maps of (a) police defined (PHIA); (b) empirically defined (EHIA) high intensity crime areas; (c) the overlay of the PHIA and EHIA maps.
Map shows COAs located in Sheffield's Basic Command Unit (BCU) K.

Comparing and combining police perceptions with police records of serious crime areas: summary of models.

Posterior means with credible intervals of parameters / Final models	Police defined HIAs	Empirically defined HIAs	Police and empirically defined HIAs
Index of ethnic heterogeneity (β_1, CI: 2.5%, 97.5%)	0.236 (0.147, 0.349)	0.019 (0.005, 0.034)	0.050 (0.039, 0.062)
No car/van (β_2, CI: 2.5%, 97.5%)	NA	0.034 (0.005, 0.066)	0.026 (0.004, 0.047)
Turnover (β_3, CI: 2.5%, 97.5%)	NA	0.109 (0.049, 0.175)	0.047 (0.005, 0.089)
Lone parent (β_4, CI: 2.5%, 97.5%)	NA	0.113 (0.056, 0.169)	0.074 (0.028, 0.120)

NA: variable did not have significant effect in the final model.

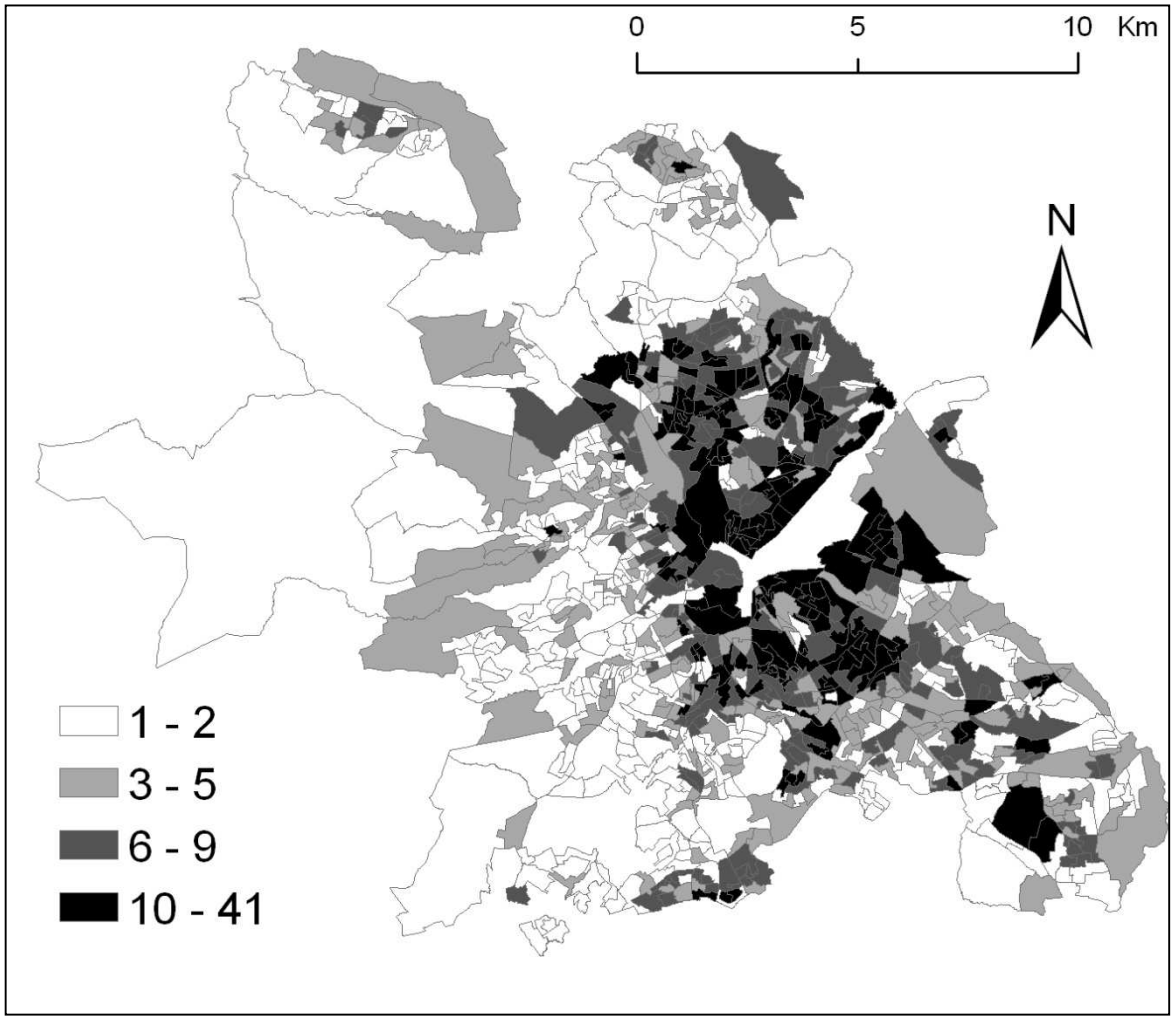
CI: credible interval

Haining and Law (2007)

Strengths and weaknesses of two different sources of information about area crime issues.

	Strengths	**Weaknesses**
Recorded crime database	- contains detail over a wide area that will probably not be known consistently by officers; - consistent procedures for recording.	- incomplete (and with a geography); - problems with locational referencing; - influenced by short term fluctuations and displacement effects.
Police perceptions	- operational knowledge - accumulated experience	influenced by - attitudes (see Rengert and Palfrey 1997); - what is/not remembered; - particular experiences.

2.3 Analysing the geographical distribution of offenders and estimating the risk of offending.



Counts of offenders, Sheffield (1995). Enumeration Districts

Geographical variation in offender rates linked to:

[a] Social processes:

- absence of effective neighbourhood social control;
- lack of social networks of association both within the community and outside it;
- inability of residents to act collectively.

Concepts: social disorganization (Shaw and McKay, 1942); low levels of social capital (Kawachi et al. 1999) and collective efficacy (Sampson et al. 1997); weakly developed networks of association (Bursik and Grasmick, 1993).

[b] Social composition:

- allocation policies of local authorities when housing problem families.

Negative Binomial GLM fitted to counts of offenders by enumeration districts: Sheffield 1995. (Haining et al. 2009).

	Negative binomial GLM
Intercept	-0.178 (0.020)
DETR* 1998 index of deprivation	0.167 (0.012)
% males 16-24 and unemployed	0.010 (0.002)
% lone parent households	0.017 (0.006)
% households in permanent dwellings rented from the local authority.	0.287 (0.093)
Adjusted pseudo R^2 (as %)	51.7%
Deviance/degrees of freedom	1.002
Estimate of the parameter of extra-Poisson variation	0.159 (0.016)
Moran score on residuals (z score)	0.116 (6.009)

* DETR Department of the Environment Transport and the Regions.
Parameter estimates with associated standard errors in brackets.

3. The challenges presented by ecological analysis.

(a) modifiable areal units problem (MAUP):

(i) scale effect (different results at different resolutions).

(ii) grouping effect (different results from different aggregations)

Both effects are linked to data smoothing

Understanding the impact of spatial aggregation on the results of data analysis is made more complex by the presence of spatial autocorrelation.

Implications for:

- mapping (pattern detection);
- hot spot detection;
- results of modelling (regression and correlation):

A summary of the effects of different types of grouping of census tracts on the correlation coefficient and regression slope parameter (Gotway and Young, 2002).

Method of grouping tracts	**Effect on correlation coefficient**	**Effect on slope parameter of simple linear regression**
Random	No systematic effect	No systematic effect
Spatial contiguity	Correlation increases to a maximum and then decreases with level of grouping	Slope parameter increases
Grouping by the dependent variable	Correlation increases with level of grouping	Slope parameter increases with level of grouping
Grouping by the independent variable	Correlation increases with level of grouping	No systematic effect

(b) Selection of “appropriate” spatial units: neighbourhoods.

(c) Incompatible spatial frameworks

(d) Areas with small populations – populations tend to be more homogeneous but statistics suffer from the small number problem.

Areas with large populations – statistics more robust (with smaller standard errors) but populations tend to be more heterogeneous

(e) Classical statistical analyses need to contend with the problems created by (inter-area) spatial autocorrelation:

(i) in dependent variable

(ii) in model residuals.

⟹ inference problems

⟹ application of geostatistics

(f) Intra-area correlation: clustering models

(g) Individual inference? Ecological bias => ecological fallacy.

Modelling the probability of offending at the individual level.

- multi-level or other forms of modelling of large data sets on individuals and combing with ecological attributes in the analysis. Expensive; confidentiality issues.

- combining ecological data with *small* amounts of individual level data to provide information on within-area covariate distributions. (Developments in spatial epidemiology: Jackson et al. 2008).

4. Final points.

-Spatial ecological analyses fit naturally into research on crime and disorder and are relevant to the way police forces operate.
- Spatial ecological analyses present a number of challenges to data analysts.
- Spatial analysis continues to be a rapidly developing area of methodological research that crime analysts ought to keep an eye on.