# Identifying the appropriate spatial resolution for the analysis of crime patterns

Nick Malleson*[1], Wouter Steenbeek†[2] and Martin Andresen‡[3]

[1]School of Geography, University of Leeds, UK
[2]Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, the Netherlands
[3]Institute for Canadian Urban Research Studies (ICURS), School of Criminology, Simon Fraser University, Burnaby, BC, Canada.

January 17, 2019

**Summary**

This research presents a new approach to estimate the most appropriate scale for the analysis of spatial point patterns. It creates a number of regular grids with iteratively smaller cell sizes and estimates the *similarity* between two realisations of a point pattern at each resolution. The method is applied to crime data from the city of Vancouver, Canada. Importantly, the results are context specific so a single 'appropriate' scale for each crime type is not identified. However, the method is nevertheless useful as a means of better estimating the appropriate spatial scale for a particular piece of analysis.

**KEYWORDS:** Spatial Scale, Spatial Similarity, Error, GISc, Point Pattern.

## 1 Introduction

A key issue in the analysis of many spatial processes is the choice of an *appropriate scale* for the analysis. For many phenomena, smaller spatial units are generally preferable because they are more likely to be homogeneous with respect to both the events under study and the population at risk, and, therefore, represent more accurately the underlying spatial pattern. As urban socio-demographics can vary considerable over quite small distances, large spatial units may hide or "smooth out" (Batty, 2005) important low-level patterns. Recognising the importance and practical benefits of (starting with) small spatial units, recent research in many social-science fields tends towards 'micro places'. This is especially true for crime science research, which is the subject of this paper.

*n.s.malleson@leeds.ac.uk
†andresen@sfu.ca
‡WSteenbeek@nscr.nl

But is there a limit to how small a spatial area should be? The aim for this work is to develop a general method that is capable of identifying the most *appropriate* spatial unit for the analysis of spatial patterns. The method is then applied to the study of various different categories of property crime in Vanouver, BC, Canada. The proposed approach adapts the multiple resolution goodness-of-fit procedure originally conceived by Costanza (1989) and combines it with a measure of spatial (dis)similarity – Andresen's $S$ (Andresen, 2009; Andresen and Malleson, 2011; Wheeler et al., 2018) – in order to identify the most appropriate scale of analysis for a particular point pattern.

## 2 Background

Recent research has found that crime concentrates at 'micro places' and, in general, a small unit of analysis is the most appropriate for many quantitative environmental criminology studies (Oberwittler and Wikström, 2009; Weisburd et al., 2009; Andresen and Malleson, 2011; Weisburd, 2015; Eck et al., 2017; Lee et al., 2017). However, care must be taken when considering micro places because of measurement issues that arise when there are more places than events (Bernasco and Steenbeek, 2017). Moreover, there is both theoretical and empirical support for considering larger spatial scales (Steenbeek and Weisburd, 2016; Schnell et al., 2017). Consequently, even if we have spatial point crime data, it can be useful to aggregate point pattern data to larger spatial units.

Additionally, determining the appropriate spatial scale is non-trivial for at least two other reasons. Firstly, high-resolution data that are required for studies at the level of micro-places, such as individual addresses, are harder to obtain than more aggregate data. Secondly, small number problems can occur when rare crime events are analysed at small spatial scales (Oberwittler and Wikström, 2009). One could, therefore, ask the question: at what point does it become unnecessary to obtain finer scale data?

## 3 Data and Methods

To answer the question of the *appropriate* spatial scale, this paper presents a new method that builds on a multiple resolution goodness-of-fit procedure (Costanza, 1989) and combines it with Andresen's S Index (Andresen, 2009; Andresen and Malleson, 2011; Wheeler et al., 2018) for measuring similarity. The aim is to find the spatial resolution that is sufficient to correctly capture the relationships that ultimately lead to the observed spatial structure of the data. The analysis source code and data (in the R language) are available, in full, at `https://github.com/nickmalleson/Multi-Scale-Error-Analysis/`.

In short, the method works by taking two realisations of the point pattern as input and placing a regular grid over them. It aggregates points to the grid and then calculates the difference between the two point datasets using Andresen's $S$. Note that if a particular cell does not have a sufficiently large expected count then there are too few points to say, with confidence, whether the two point patterns are similar or dissimilar in that cell. In these cases the cell is removed from the analysis and has no influence on the global similarity measure.

The resolution of the grid is then increased by adding one row and one column of cells, shrinking the grid cells' size so that the grid covers the point data again, re-aggregating the points to the grid, and then re-calculating the goodness-of-fit. This process is repeated a number of times to allow a comparison of the similarity of the point patterns at the different resolutions in the form of a graph. Fig 1 broadly illustrates the method.

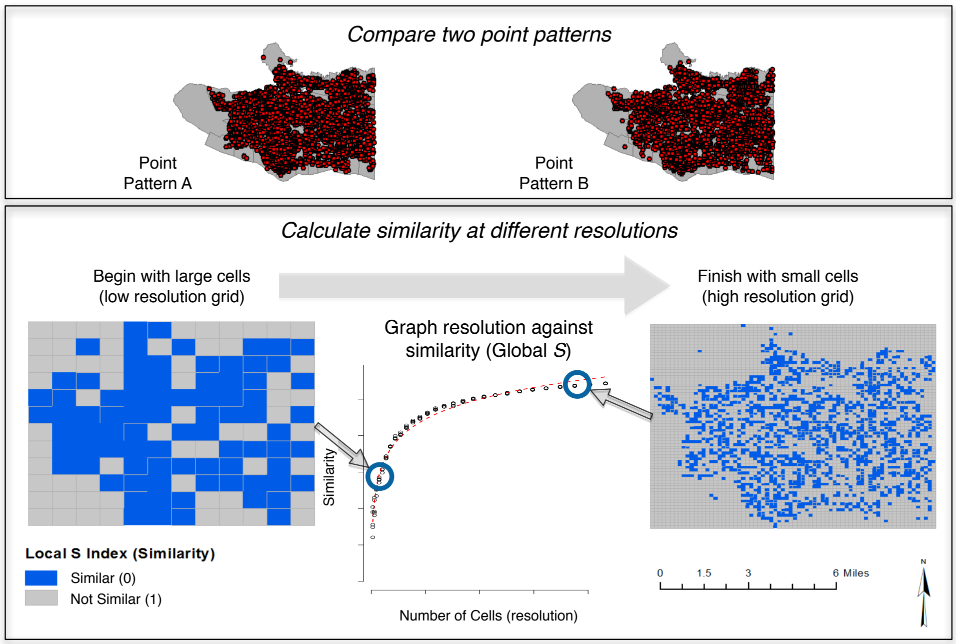


Figure 1: A broad overview of the method proposed here.

The data used to test the method represent calls or complaints made to the Vancouver Police Department. They are publicly available through the City of Vancouver Open Data Catalogue[1]. Two realisations of each point pattern (crime events in this case) are required for the analysis. The chosen realisations need to be broadly similar so that differences between them are a result of the choice of the resolution, not an artefact of a difference in the underlying process that produced them. Therefore the individual crime datasets are divided into two by the year that crimes took place; 2015 events are compared to 2016 events. The following four crime categories have been chosen:

1. Breaking and entering – residential (BNER);
2. Breaking and entering – commercial (BNEC);
3. Theft from vehicle (TFV);
4. Theft of bike (TOB).

[1] https://data.vancouver.ca/datacatalogue/crime-data.htm

## 4 Results

Fig 2 illustrates how the mean similarity changes with resolution. With all crime types, the mean similarity initially increases. After reaching a peak, it then remains relatively consistent or (as in the case of residential burglary) decreases.

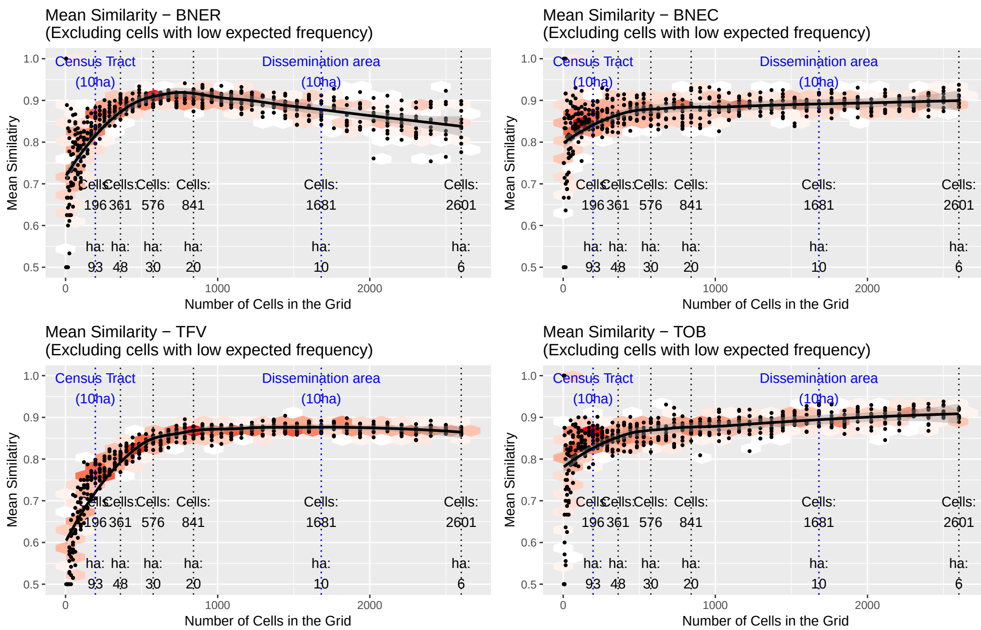


Figure 2: A scatter plot of the global similarity ($S_g$) of all cells at each resolution.

## 5 Discussion

Using the graphs in Figure 2 to identify the most appropriate scale is non trivial. As the size of the spatial units changes, there are two distinct statistical processes that take place. These are illustrated in Fig 3. The pertinent question therefore becomes: *where, on the graph, does the most appropriate spatial scale lie?* Interestingly, the behaviour of the global similarity varies by the crime type in question, and therefore the explanation needs to be tailored for each one.

- **Residential Burglary (BNER).** With BNER, there is a peak in similarity when cells are approximately 20 hectares in size. This size, which is close to twice the square area of a dissemination area (a Canadian census neighbourhood that is similar to an Output Area) appears to be the most appropriate spatial scale at which to analyse the phenomenon. Of course, using larger units will hide lower-level patterns so generally smaller is better, but these results suggest that when using units that are much smaller than 20 hectares the *noise*, that will always be present with such spatial phenomena, is captured instead of the *signal*.

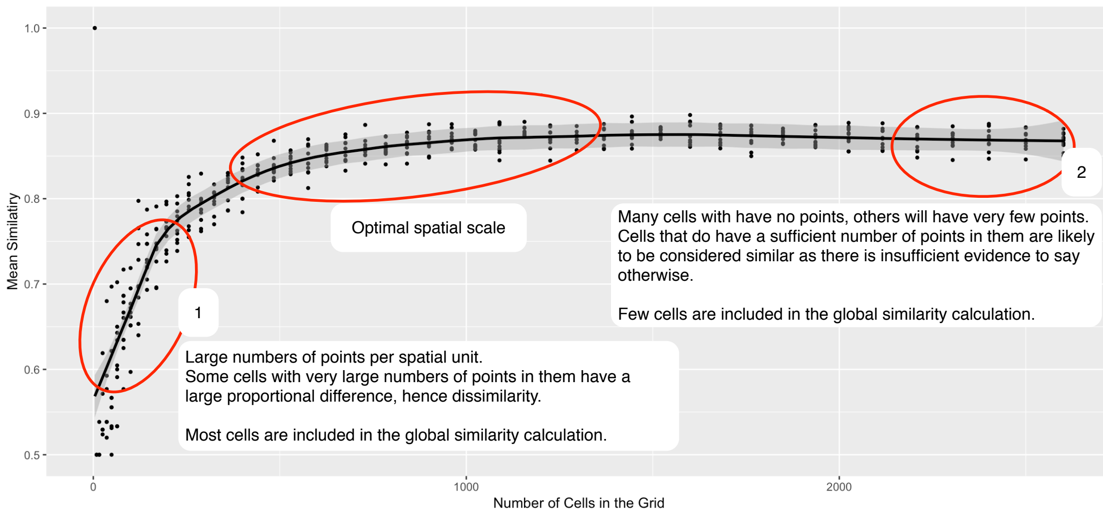


Figure 3: An explanation of the change in similarity with the size of the spatial unit.

- **Theft from Vehicle (TFV)**. In this case, the similarity increases until the spatial units are approximately 20ha and then plateaus. With TFV, therefore, there appears to be no *disadvantage*, in terms of signal v.s. noise, to using spatial units that are smaller than 20ha.
- **Commercial Burglary (BNEC)** and **Theft of Bike (TOB)**. As with BNER and TFV, the similarity increases rapidly initially but then, rather than plateauing, it continues to increase gradually. This suggests that, even with cells of only 6ha, there is still a signal that can be identified above the background noise.

A further important consideration is the number of points in a dataset and, similarly, the point density. As the number of points increases then, if everything else remains equal, the appropriate spatial scale becomes finer. This is because with two *similar* patterns, more points means that the cell size will be able to become quite small before noise begins to hide the signal. This is important because phenomena that occur frequently, or at least frequently in a specific area (e.g. a highly clustered phenomena) can be reliably disaggregated to quite small cells, at least in the areas where the phenomena occurs densely. The 'appropriate spatial scale' not only depends on the abundance of the phenomena itself, but also the degree to which it is clustered.

## 6 Conclusions

This paper has presented a new approach that can be used to identify the most *appropriate scale* for the analysis of spatial point processes. The method is dependent on both the size of the point patterns (the number of events) and the degree of clustering, so it is doubtful that a single 'appropriate' scale will ever be identified for a phenomenon. But the method is nevertheless useful as a means of better estimating what spatial scale might be appropriate for a particular piece of analysis.

## 7 Acknowledgements

This work has been funded by the Economic and Social Research Council (ESRC) grant numbers ES/L009900/1 and ES/L011891/1, awarded to NM as principal-investigator and co-investigator respectively. The work was also made possible through the award of a Research Fellowship from the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR).

## References

Andresen, M. and N. Malleson (2011). Testing the Stability of Crime Patterns: Implications for Theory and Policy. *Journal of Research in Crime and Delinquency 48*(1), 58–82.

Andresen, M. A. (2009). Testing for similarity in area-based spatial patterns: A nonparametric Monte Carlo approach. *Applied Geography 29*(3), 333–345.

Batty, M. (2005). Agents, cells, and cities: New representational models for simulating multiscale urban dynamics. *Environment and Planning A 37*, 1373–1394.

Bernasco, W. and W. Steenbeek (2017). More Places than Crimes: Implications for Evaluating the Law of Crime Concentration at Place. *Journal of Quantitative Criminology 33*(3), 451–467.

Costanza, R. (1989). Model goodness of fit: A multiple resolution procedure. *Ecological Modelling 47*(3-4), 199–215.

Eck, J. E., Y. Lee, S. O, and N. Martinez (2017). Compared to what? Estimating the relative concentration of crime at places using systematic and other reviews. *Crime Science 6*(1).

Lee, Y., J. E. Eck, S. O, and N. N. Martinez (2017). How concentrated is crime at places? A systematic review from 1970 to 2015. *Crime Science 6*(1).

Oberwittler, D. and P.-O. H. Wikström (2009). Why Small Is Better: Advancing the Study of the Role of Behavioral Contexts in Crime Causation. In D. Weisburd, W. Bernasco, and G. J. Bruinsma (Eds.), *Putting Crime in Its Place*, pp. 35–59. New York, NY: Springer New York.

Schnell, C., A. A. Braga, and E. L. Piza (2017). The Influence of Community Areas, Neighborhood Clusters, and Street Segments on the Spatial Variability of Violent Crime in Chicago. *Journal of Quantitative Criminology 33*(3), 469–496.

Steenbeek, W. and D. Weisburd (2016). Where the Action is in Crime? An Examination of Variability of Crime Across Different Spatial Units in The Hague, 2001–2009. *Journal of Quantitative Criminology 32*(3), 449–469.

Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology 53*(2), 133–157.

Weisburd, D., W. Bernasco, and G. Bruinsma (Eds.) (2009). *Putting Crime in Its Place: Units of Analysis in Geographic Criminology.* New York: Springer. OCLC: ocn233934836.

Wheeler, A. P., W. Steenbeek, and M. A. Andresen (2018). Testing for similarity in area-based spatial patterns: Alternative methods to Andresen's spatial point pattern test. *Transactions in GIS 22*(3), 760–774.