ARTICLE IN PRESS

Computers, Environment and Urban Systems xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

Analysis of crime patterns through the integration of an agent-based model and a population microsimulation

Nick Malleson *, Mark Birkin

School of Geography, University of Leeds, Leeds LS2 9JT, United Kingdom

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Agent-based modelling
Crime simulation
Burglary
Microsimulation

ABSTRACT

In recent years, criminologists have become interested in understanding crime variations at progressively finer spatial scales, right down to individual streets or even houses. To model at these fine spatial scales, and to better account for the dynamics of the crime system, agent-based models of crime are emerging. Generally, these have been more successful in representing the behaviour of criminals than their victims. In this paper it is suggested that individual representations of criminal behaviour can be enhanced by combining them with models of the criminal environment which are specified at a similar scale. In the case of burglary this means the identification of individual households as targets. We will show how this can be achieved using the complementary technique of microsimulation. The work is significant because it allows agent-based models of crime to be refined geographically (to allow, for example, individual households with varying wealth or occupancy measures) and leads to the identification of the characteristics of individual victims.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

An early contribution of environmental criminology has been to illustrate systematic variations in the profile of criminal activity between different area types, such as persistently lower levels of crime in rural communities than intensely urban neighbourhoods. In recent years, however, criminologists have become interested in understanding variations at progressively finer spatial scales, right down to individual streets or even houses. The analysis of crime at such a fine scale is also supported by recent developments in computer simulation such as agent-based modelling. In earlier work, the value in representing criminals and their behaviour as individuals has been demonstrated within a richly specified modelling framework. Ultimately such models exploit the fact that aggregate crime patterns are no more or less than the sum of a series of unique events, each bringing together a criminal and a victim in space.

To date, agent-based models of crime have been more successful in representing the behaviour of criminals than their victims. In a sense these models are hybrids which combine individual criminal actors with a less disaggregate view of the environments in which they operate.[1] In this paper it is suggested that individual representations of criminal behaviour can be enhanced by combining them with models of the criminal environment which are specified at a similar scale. In the case of burglary this means the identification of individual households as targets. We will show how this can be achieved for an agent-based model using the complementary technique of microsimulation.

The work is significant for a number of reasons. It allows agent-based models of crime to be refined to allow for the variable attractiveness of specific targets, for example households with high wealth or low occupancy. Second, by identifying individual victims we allow the possibility of including repeat victimisation itself as a major contributor to crime patterns. Studies have shown that repeat victimisation is the strongest at risk factor for the victims of burglary (Tseloni, 2006). We introduce the possibility of further disaggregation of victim characteristics and behaviours, such as the influence of age, ethnicity or household composition. Finally, we hope that this work will also be of interest to those who are looking to disaggregate models of individual behaviours in other sectors such as retailing, health or education.

In Section 2 of the paper the importance of spatial environments for crime modelling will be discussed. The individual level modelling techniques of agent-based modelling and microsimulation are reviewed. The way in which individual-based models have been implemented in the context of crime is described in Section 3, together with a discussion of the means for model validation. A method for integrating individual models of both criminal and victim is presented in Section 4 of the paper, before a discussion of some numerical experiments and results from the new model in

* Corresponding author. Tel.: +44 113 343 6757.
E-mail address: n.malleson06@leeds.ac.uk (N. Malleson).

[1] There is some similarity here to retail models in which the individual behaviour of retailer agents has been combined with a more coarsely-grained view of customer activity (Heppenstall et al., 2007).

0198-9715/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compenvurbsys.2012.04.003

Please cite this article in press as: Malleson, N., & Birkin, M. Analysis of crime patterns through the integration of an agent-based model and a population microsimulation. *Computers, Environment and Urban Systems* (2012), http://dx.doi.org/10.1016/j.compenvurbsys.2012.04.003

Section 5. The paper concludes with some reflections, conclusions, and suggestions of the most immediate priorities for further work.

2. Background – modelling crime

Crime is inherently a human phenomenon; a single crime event is the result of the motivations and behaviour of the criminal, victim and other people who might be able to influence the event (Cohen and Felson, 1979) as well as their relationships with or attitudes to the surrounding environment (Brantingham and Brantingham, 1993). These complex human factors, coupled with vast environmental complexity, make crime very difficult to understand, predict and model. However, crime does not occur at random and a considerable body of literature has been developed in order to identify the underlying drivers required to model the 'crime system'. This section will discuss the relevant theoretical and practical approaches for understanding and modelling crime. It will be shown how the individual-level modelling techniques employed – agent-based modelling and microsimulation – are ideally suited to capturing the dynamics of the crime system and offer additional insight through their integration.

2.1. Environmental criminology and 'traditional' crime modelling

Although the earliest examples of spatial crime analysis date back to the 18th century (e.g. Glyde, 1856), the term "environmental criminology" was not coined until 1971, when Jeffery (1971) called for the development of a new school to focus on the environment in which crime occurs (Andresen, 2009). As research has progressed, environmental criminology studies have focussed on the effects of the environment at progressively smaller scales, to the extent that "crime at places" (Eck and Weisburd, 1995) research now generally concentrates on individual streets or houses. This progression in quantitative research has been led by corresponding theoretical developments; the major theories associated with environmental criminology (Brantingham and Brantingham, 1981; Cohen and Felson, 1979; Clarke and Cornish, 1985) focus on the spatio-temporal behaviour of individual people and their immediate surrounding environment.

Models of crime have also followed the trend towards higher resolution geographies. The tradition of using spatially aggregated census data is being replaced by work that models at the level of the individual street (Johnson and Bowers, 2009) and house (Tseloni, 2006). In terms of methodologies employed, the 'traditional' regression approaches (multivariate, poisson, negative binomial and logistic, etc.) are being advanced through the use of techniques that are common in other disciplines – such as a discrete spatial choice model to study target choice (Bernasco, 2004) and multi-level modelling to examine property crimes (Tseloni, 2006).

Regardless of the precise method employed, most crime models are generally linear. The crime system, on the other hand, is a *complex system*; it is made up of numerous interacting elements, exhibits non-linear behaviour and involves feedback. Although linear models are "computationally convenient", they cannot capture the dynamics of such systems (Eck and Liu, 2008). Furthermore, statistical models aim to reduce the number of explanatory variables which can make it more difficult to account for environmental complexity and the human–human or human–environment interactions that drive the system. Similarly, spatial realism is often compromised through the use of simple Euclidean distance measures that do not capture the richness of the physical environment (such as the presences of roads, parks, and rivers). These factors mean that although linear crime models have proven to be an essential tool for crime analysis they are flawed in terms of capturing the underlying dynamics that drive the system. Individual-level models, on the other hand, focus on manipulating the individual units that drive the system – in this case criminals, victims, managers, households, etc. – and are thus much better suited to modelling the dynamics of complex systems.

2.2. Agent-based modelling (ABM)

In general, the central drawback to statistical crime models is that they are not able to capture the underlying processes that drive the crime system – one that is characterised by the behaviour of individual people who have their own unique psychology and interact in a rich social and physical environment. An alternative approach to modelling these types of systems, as opposed to controlling them from the 'top down' with an equation, is to simulate the behaviour of the individual actors that drive the system directly. This is the approach taken with ABM. Unlike statistical models, an agent-based model is comprised of individual entities called *agents*, who are able to behave autonomously. Agents are placed in a virtual environment and they are able to interact with each other and with the environment. A model is executed over a number of iterations and at each iteration the agents have the ability to assess their situation and make a decision about their future actions. Realistic human behaviour can be built into the model through the means that the agents make their decisions – behavioural complexity can range from simple rule-based systems, (e.g. Schelling, 1969), through to advanced cognitive frameworks (Schmidt, 2000). Overall, creating models in this manner from the "bottom-up" (Epstein and Axtell, 1996) is a much more natural means of describing complex systems (Bonabeau, 2002).

Although there is great potential to be offered by agent-based models, there are inevitably some drawbacks. The advantage of being able to model human behaviour could itself be construed as a drawback because it is actually extremely difficult to simulate human psychology in a computer model. This encourages models to have minimal behavioural complexity (O'Sullivan and Haklay, 2000) which is not necessarily justified. Also, computation time is often a problem for models: their probabilistic nature means that they must often be run numerous times and the time required to process each individual agent inevitably increases with behavioural complexity. Fortunately, there are efforts to make high-performance computer hardware more readily available which can alleviate some of the computational problems.

Models must also face the *equifinality* problem, which is where many models might match a single set of calibration data. Therefore the number of variables that can be used in a model are limited where sufficient data are not available. Because agent-based models often work at the level of the individual person or household, obtaining large amounts of high-quality calibration data can be problematic. Similarly, data is often required in order to characterise the agents that make up the system or to describe aspects of the social or physical environment. For example, with the burglary model employed by this research (which is discussed in Section 3.2) it is necessary to create a virtual representation of all the potential victims of burglary (i.e. all households in the study area). Although the 2001 UK census provides sufficient socio-demographic information to describe *neighbourhoods*, a mechanism is required to identify or estimate the individual households themselves. Without these micro-level datasets it can be extremely difficult to initialise and to validate agent-based models. Fortunately, the technique of *microsimulation* can be used for this task.

2.3. Microsimulation

Microsimulation is a comparable technique to agent-based modelling because it also represents a population as a set of distinct entities rather than by groups. Typically it is seen as a means for applying well-defined rules to a wide variety of individual circumstances in order to achieve insights with real predictive or applied value. For example, microsimulation can be used to simulate processes such as birth, death, and migration at the level of the individual household to estimate household-level population change over time (Wu et al., 2008). Although there is no clear distinction between microsimulation and ABM, generally agent-based approaches focus on richer behavioural models and on the interactions between individuals and their environment whereas microsimulation is more suited to situations with clearly defined transition rules (Wu et al., 2010).

As well as running individual-level simulations, microsimulation can be used as a means of disaggregating data. For example, microsimulation has been used to disaggregate the British Crime Survey and simulate the effects of various policy decision on the local populations who they were targeted at (Kongmuang, 2006). A commonly used source for models in the UK, which is unparalleled in its robustness and scope, is the decennial census. Although census data are released at relatively small geographical areas (an 'output area' usually contains only 100 houses) this cannot be used to seed an agent-based model unless it can be further disaggregated to the level of the individual household or person. Therefore, the coupling of an agent-based and microsimulation model offers considerable advantages. The following sections will outline the two models used here in more detail before identifying how they have been integrated and the benefits of doing so.

3. The population reconstruction and burglary models

It has been shown that in order to create accurate predictive/explanatory models of crime there is a substantial benefit to using a fine-scale geography (e.g. the level of the individual house) and to simulating the individual actors that are responsible for generating higher-level crime patterns. Hence the coupling of agent-based and microsimulation models offers modelling advantages on two fronts: using microsimulation it is possible to create high-quality individual-level data to characterise the actors; and with agent-based modelling it is possible to create realistic behavioural rules and an accurate virtual environment in which to simulate their behaviour. This section will introduce the two models employed in this research to explore the characteristics of burglary victims. The Population Reconstruction Model (Birkin et al., 2006) that can be used to disaggregate census data and the *BurgdSim* model (Malleson et al., 2012) which is an advanced agent-based model of residential burglary.

3.1. The Population Reconstruction Model (PRM)

Although it is robust, comprehensive and accurate, the UK census fails to provide a spatially disaggregate representation of individual people and households. This high-resolution representation of individuals is essential for the modelling of complex systems such as crime. Therefore a microsimulation program called the Population Reconstruction Model (PRM) has been developed to use a combination of census Small Area Statistics (socio-demographic data released at the lowest level of spatial disaggregation) and the Sample of Anonymised Records (a set of anonymous, a-spatial individual census records) to provide synthetic lists of the entire population of any city or region in the country.

The PRM (Birkin et al., 2006) uses an iterative reweighting procedure to allocate synthetic households to small areas, using attributes ranging from age, marital status, ethnicity and gender to occupation and health, housing tenure and household composition. Each characteristic is weighted to the neighbourhood of any small area which is to be reconstructed. For example, in a multi-cultural area, ethnic minority groups will attract an increased weighting; in areas of social housing then privately owned accommodation will attract a reduced weighting; and so on. A variety of microsimulation techniques are applicable to the problem of synthetic reconstruction (Williamson et al., 1998).

Determining whether or not the synthetic population is accurate is non-trivial because there are no data that can be used to validate it directly (if individual-level data were available in the first place the PRM procedure would be unnecessary). The most common method of assessing validity is to compare the aggregate synthetic population to the original Small Area Statistics under the assumption that if the aggregate populations correspond then the synthetic population is a close representation of the real population. To this end, it has been shown that the PRM outputs have an extremely close match to the small area distributions from which they are derived and hence the PRM is accurate (Harland et al., 2012).

Table 1 illustrates the personal and household attributes that are currently available as output from the PRM as these will be used to characterise the individuals in the agent-based model.

3.2. The BurgdSim model

3.2.1. The burglar agents

Environmental criminologists have emphasised the importance of addressing the intricacies of the physical or social environment (Brantingham and Brantingham, 1993; Eck and Weisburd, 1995) and the effects of individual peoples' behaviour (Cohen and Felson, 1979; Clarke and Cornish, 1985) in order to construct accurate models of crime. The *BurgdSim* crime model is an advanced agent-based model of residential burglary that aims to capture these elements. Offenders in the model (virtual burglars) are represented as individual agents, who are able to navigate a realistic urban environment performing normal day-to-day behaviours. In its current form, these behaviours include *sleeping*, *socialising* and *using substances* which, although obviously a vast simplification on real human behaviour, have been identified as being the most important drivers for many burglars (Wright and Decker, 1996; Cromwell et al., 1991; Wiles and Costello, 2000).

To control the agents, the PECS (Physical Conditions, Emotional State, Cognitive Capabilities and Social Status) artificial intelligence

Table 1
The individual and household attributes that are contained in the synthetic population output by the PRM.

Attribute	Description
House size	The number of people who live in the household
House type	The type of the house building. Can be one of: *detached*, *semi-detached*, *terraced* or *flats*
Age	The age of the individual in single year groups
Gender	The gender of the individual
Ethnicity	The person's ethnicity. The census groups are aggregated, for simplicity, into the following four categories: *white*, *asian*, *black*, *other*
Marital Status	Whether the person is married or unmarried
Employment Status	The person's employment. As with ethnicity, the census employment categories are grouped into one of: *managerial*, *intermediate*, *manual* or *other*. Employment status is also later used as a proxy for socioeconomic status

framework (Schmidt, 2000) was used to equip the agents with realistic, dynamic behaviour (Malleson et al., 2010). With PECS, an agent's behaviour is determined by the strength of different needs (socialising, using substances or sleeping in this case). It is the strongest of these needs that determines their current behaviour at any point in time. Committing burglary is a response to having to meet a need that requires money – socialising or committing burglary – because in the model the agents cannot gain wealth through legitimate employment. In this manner it is possible to build up city-wide burglary patterns by simulating the behaviour of the individuals who are ultimately responsible for the individual crimes. It is possible to create heterogeneous agent behaviour by varying key parameters that determine where agents will start looking for burglary targets and which houses they find the most attractive. For example, it would be possible to create a "professional" who was comfortable travelling larger distances than other agents in search of more lucrative targets. Varying these behaviours will have a substantial impact on the model outcomes, but a full exploration of heterogeneous offender behaviour and its effect on city-wide burglary patterns must be left for future work. For more information about behaviour validation, the interested reader can refer to Malleson et al. (2012).

Each agent is assigned a house as their home and they start the simulation there with low needs (i.e. they are satisfied and are not motivated to perform an action). Over time their needs increase and they become motivated to attempt to satisfy them. The simulation is configured so that on a typical day, an agent must sleep for 8 h, socialise for two hours and purchase drugs once. The income gained from a single burglary is set to a constant amount which is sufficient to allow the agent to purchase drugs once and socialise for 2 h (hence, on average, an agent will need to burgle once per day). There is no law enforcement in the model, so agents base their burglary decision purely on their own internal needs and the attributes of the surrounding environment. However, on some days the agent might not find a suitable target which will lead to them become more desperate and commit multiple burglaries on a later day. In this sense the agents are truly autonomous; the amount of time they spend performing different activities depends entirely on their own behaviour, there is no central control. See Malleson et al. (2012) for more details about the burglar agents.

Another important agent characteristic are their *cognitive maps*. Agents do not have global knowledge of their environment and instead they build up their awareness of houses as they pass them on routine travels. For example, an agent might become aware of a potential burglary target whilst on the way to socialise and later return to burgle the house. This is a powerful component of the model because it brings it much more closely in line with criminology theory (Brantingham and Brantingham, 1981; Clarke and Cornish, 1985) and means that the urban form of an area will have an influence on burglary patterns (houses that are situated in areas that the burglar agents are unlikely to have passed through will have lower risk).

3.2.2. The virtual environment

As with the representation of the agents, the virtual environment in the model has been designed to be as realistic as is necessary for a burglary simulation. To this end there are three distinct layers that make up the environment:

- The *buildings* layer contains physical buildings. These are the houses that the burglar agents can attempt to burgle and have been generated from MasterMap geographic data as depicted in Fig. 1. Each house is a unique object with different physical attributes that reflect current theoretical understanding of the crime system, e.g. ease of access to the house, its visibility to neighbours, etc.
- The *transport* layer is another physical layer and it makes up the transport network for the simulation area. Again using MasterMap data, there are distinct geographical objects to represent roads, rail networks and bus routes. Roads also have attributes that determine whether or not they are car or pedestrian accessible. Realistic routing behaviour is obtained by varying the speed that agents can drive along roads so that agents with cars are encouraged to drive on major roads rather than using minor ones.
- The remaining layer is the *community* layer which, unlike the buildings and transport layers, is used to account for the effects that *other people* will have on a potential crime occurrence. For example, high levels of community cohesion have been linked to low levels of violent crime because local people are more likely to intervene to prevent a crime from occurring (Sampson et al., 1997). Similarly, areas with large numbers of residents who are at home during the day can offer informal protection and reduce the burglary risk.

3.2.3. The need for a synthetic population

In terms of physical attributes, the model virtual environment is highly detailed and high resolution; it is able to represent individual roads and houses which modern environmental criminology research suggests are important. However, there are some drawbacks with the community layer that exist due to the absence of household-level population data. Currently, the layer includes (among others) the following two attributes:

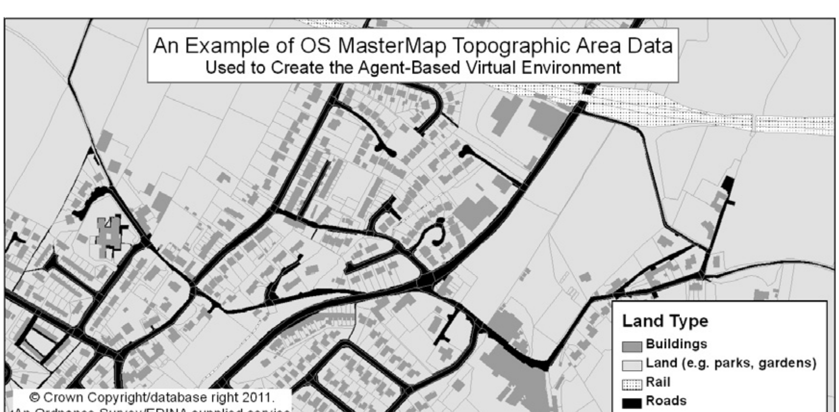

Fig. 1. An example of the Ordnance Survey MasterMap data that are used to create the virtual environment.

Please cite this article in press as: Malleson, N., & Birkin, M. Analysis of crime patterns through the integration of an agent-based model and a population microsimulation. *Computers, Environment and Urban Systems* (2012), http://dx.doi.org/10.1016/j.compenvurbsys.2012.04.003

- **Occupancy**. An estimate for whether or not a household is occupied at a particular time based on the employment statuses of the people who live in the area from the UK census. For example, student houses are more likely to be occupied during the day.
- **Attractiveness**. A measure of the affluence of houses in the area, also based on census data.

Although some factors (such as community cohesion) correspond to communities rather than individuals and should therefore be included at an aggregate level, the *occupancy* and *attractiveness* variables will not necessarily be homogeneous across an entire community. It would clearly be preferable to model them at the household level. Furthermore, environmental criminology research has shown that victim behaviour is an important determinant of household burglary risk so it is a major drawback that a model with such an accurate representation of the physical environment must aggregate certain key variables due to a lack of household-level demographic data.

Therefore this research will take advantage of the PRM microsimulation model to create estimates of *occupancy* and *attractiveness* for every individual household in the simulation area, rather than assuming all houses in an area are identical in this respect. Furthermore, by attaching additional information about the synthetic individuals to households (such as the residents' gender, and ethnicity) the research is able to perform illuminating post-simulation analysis of the victims of crime.

It should be noted that victims are still not represented as agents; each household has heterogeneous levels of vulnerability but at this stage households do not 'think' in the way that burglar agents do (they will not react to a burglary). This is an obvious avenue for future research as it has been shown to have advantages in other work (Malleson et al., 2010). Nevertheless, the integration of the *BurgdSim* agent-based model with the PRM microsimulation model offers considerable advantages for the burglary simulation in terms of bringing the existing model in-line with current criminological thinking.

3.3. Validating the simulation results

Before validating the simulation results (by comparing model results to known data) it is necessary to verify that the model is logically consistent – a process often termed "verification" (Castle & Crooks, 2006) or "inner validity" (Axelrod, 1997). To verify that the model had been implemented correctly, it was executed in three different types of environment: a 'null' environment in which each agent's journeys took a set amount of time; a 'grid' environment in which roads and houses were situated on a regular grid; and finally a realistic 'GIS' environment that closely represented the real area under study. By varying the environmental complexity in this manner it was possible to ensure that changes in controlling factors had the expected influences on model outcomes in the absence of a complex, confounding geography. For full details and results of verification experiments, see Malleson, Heppenstall, Evans, and See (2010).

After verification, the model can be compared to observed data to determine how closely it reflects known system conditions. As with the validation of microsimulation models, validating agent-based models is a divisive subject as there is no established method that can be used across different research projects. The *BurgdSim* model was calibrated and validated by comparing the model's output burglaries to known burglary data provided by the police. The model is stochastic so, during the process of calibration, it was run a sufficient number of times (usually 50–100) to ensure that the aggregate results were consistent (Malleson, 2010).

The process of *comparing* the simulated data to the expected data is non-trivial because the two datasets are made up of points in space. Therefore there are a multitude of ways to answer the question "how similar are these two point patterns". A common approach is to spatially aggregate the point data to some administrative boundary and then apply traditional goodness-of-fit statistic such as R^2 or the Standardised Root Mean Square Error (SRMSE). However, aggregation to administratively-defined areas makes the approach highly susceptible to the modifiable areal unit problem (Openshaw, 1984).

To avoid these drawbacks here, a new method was developed, following Costanza (1989), to assess the difference between two point datasets. The method, which was first published in Malleson (2010), takes advantage of traditional goodness-of-fit statistics, but instead of aggregating to an administrative boundary it places a number of cellular grids of varying resolutions over the study area and counts the number of points in each grid cell. By using various grids the method is able to minimise the effects of the modifiable areal unit problem. Also, it is possible to give local estimates of difference using the *relative percentage difference*. For two cells, y_i and y'_i, this is defined as the difference between the proportions that the cells contribute to the total observation count:

$$100 * \left(\frac{y_i}{\sum y}\right) - 100 * \left(\frac{y'_i}{\sum y'}\right) \quad (1)$$

The advantage with using this method to calculate cell difference is that it is not influenced by the total number of points in the datasets. The results presented in Section 5 use this method to explore spatial differences.

4. Integrating the models

4.1. Data preparation

The first stage in the process of integration is preparing the data for input into the *BurgdSim* model. The model represents the environment with Ordnance Survey MasterMap Topographic Area data

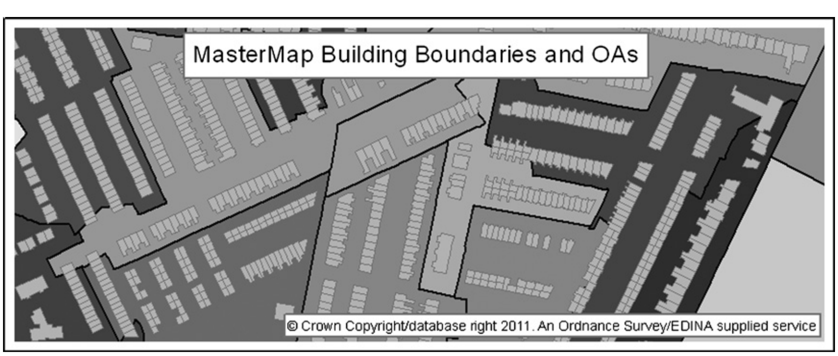

Fig. 2. Buildings as generated from Ordnance Survey MasterMap Topographic Area data and their associated output area boundaries.

Table 2
Household occupancy behaviour as implemented in the model.

Group	Description	*P* (house occupied)
Family	The house has young children and someone will be at home during the day to look after them	Higher probability of the house being occupied during the day and in the evenings
Students	The household is made up of university students	Higher probability during the day but less in the evenings as the students socialise
Unemployed	No one in the household is employed	Higher probability of the house being occupied at all times

which is a vector GIS dataset containing the individual boundaries of buildings. The PRM model, however, uses census data that are published at the output area (OA) geography and hence the synthetic population created are spatially referenced to their associated OA, not to an *individual house* within an OA. Therefore the main challenge in terms of data preparation is disaggregating the synthetic population to the household level. Fortunately, the PRM is able to estimate the type of house that a synthetic household lives in (detached, semi-detached, terraced or flats) and this information is used to assign synthetic households to buildings. At present, each household is randomly assigned to a building of the correct type within the target OA. For illustrative purposes, Fig. 2 contrasts the OA and building geographies.

4.2. Adapting the burglary model

The 'key' household burglary risk variables that could not be established without the integration of a microsimulation model are a measure of household affluence and the likelihood of the house being occupied. The *affluence* attribute is estimated from the employment of the head of household which can be one of four types: *managerial*, *intermediate*, *manual* and *other* (including unemployed, retired and students). Household affluence is estimated directly with *managerial* types being the most affluent and *other* the least.

In terms of estimating occupancy, to coincide with the *BurgdSim* model it is necessary to place each household into one of the groups illustrated in Table 2 (note that if a household does not fall into one of the specified groups then they are assumed to have typical daytime jobs). This can be accomplished by examining the employment type of the head of household as well as the other people who live in the household. For example, if the household contains young children it is assumed to be of the 'family' type. Estimating students and unemployed people is less straightforward, as Section 4.3 will discuss. It should be noted that occupancy is a probability rather than a binary value. For example it is more probable that a house containing unemployed synthetic people will be occupied during the day than one where all residents work. When burglar agents make their decision about whether or not to burgle, this probability is considered along with other variables such as the apparent security of the house, the volume of pedestrian/vehicle traffic on the adjacent road, and the visibility of the house to neighbours.

Although occupancy and attractiveness are the only two household-level variables that the agent-based model requires, the microsimulation also provides a range of person- and household-level factors (these were outlined in Table 1). We will show that although these factors do not influence the outcome of the simulation, their post-simulation analysis is illuminating.

4.3. Drawbacks with the integration approach

Sections 3.1 and 3.2 explained that the PRM and *BurgdSim* models have been thoroughly tested and calibrated and hence will

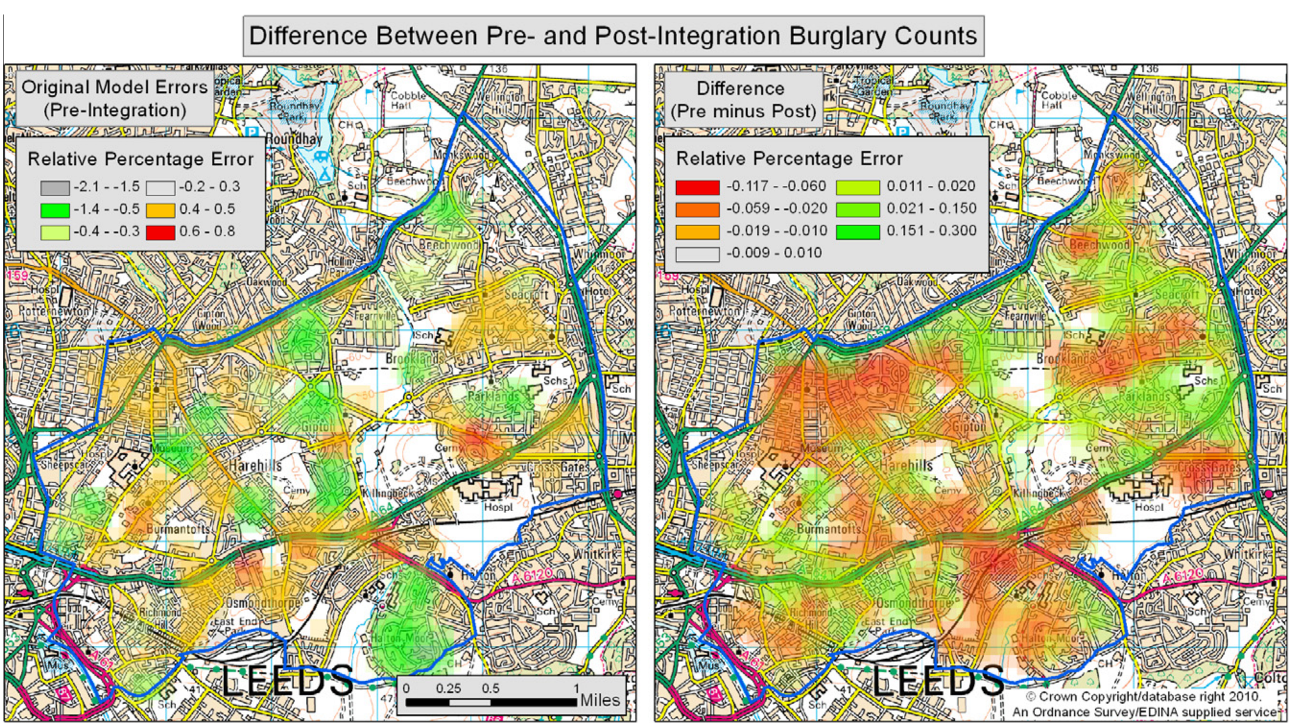

Fig. 3. A comparison of original (pre-integration) model errors (left) with the errors pro- duced by deducting post-integration burglary counts from the original burglary counts (right). It is clear that although there are some differences between the pre- and post-integration models, the differences are very small in comparison with the original model errors.

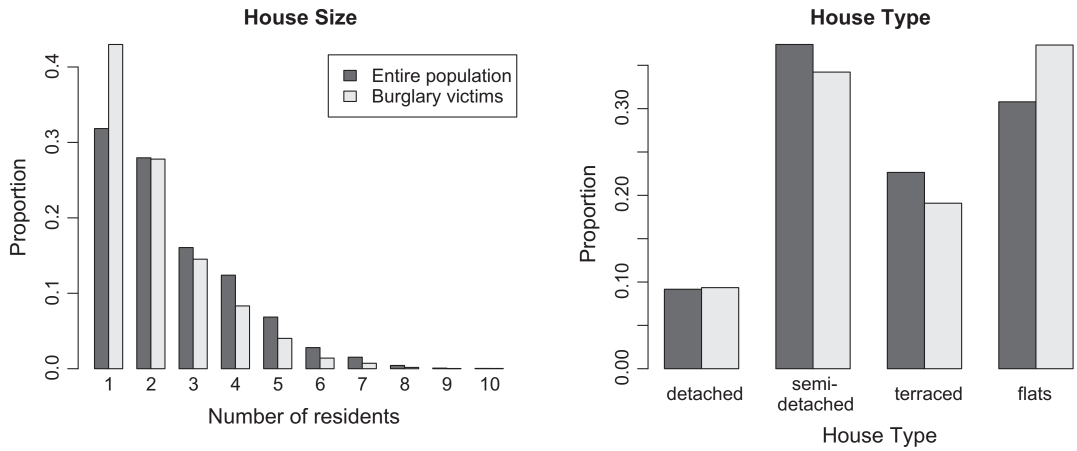

Fig. 4. Proportions of household attributes for all houses in the simulation area and the subset of burglary victims.

produce minimal error. However, there are clearly a number of ways in which error can arise in the data preparation stages and immediate future work will explore how this process can be improved. The first drawback relates to disaggregating the synthetic population and the most obvious means of improving this would be to base the allocation of households to houses on more than simply house type. For example, the affluence or income of the synthetic household could be used to assign richer families to physically larger houses or those that are more expensive (assuming house price data are available).

The other drawback comes with estimating the type of the household in terms of occupancy. Although it is relatively simple to estimate families, there are no attributes in the population currently output by the PRM that can be used to determine whether the household is made up predominantly of *students* or *unemployed* and no means of estimating *part time workers* (which is another attribute that the burglar agents can assess). Presently, it is assumed that if the head of the household is part of the *other* employment group then they are unemployed unless their ages are between 18 and 24 in which case they are a student. Therefore another obvious means of improving the integration process would be to generate a synthetic population with a richer set of attributes to represent employment type and income.

Although there are drawbacks with the process of integration, these are ameliorated because (as Section 5 will show) the use of individual victim data does not substantially influence the *aggregate* burglary patterns. Therefore they do not impact on the important insights from the model that can be gained from an assessment of the characteristics of individual burglary victims. The reason for the similarity in the aggregate patterns is largely because each output area is relatively small (the mean square area of OAs within the simulation boundary is 0.02 km^2). Therefore it is extremely likely that an offender agent will be aware of a vulnerable target even if it should, in reality, be located in a different building somewhere else in the output area. Hence the benefits of integrating the two models here are in terms of assessing which people have become victims, rather than accurately estimating in which houses the victimised people actually reside.

5. Experiments and results

5.1. The modelling scenario

The chosen scenario area is part of the city of Leeds, UK. In particular, an area of approximately 1700 hectares located to the east of the city centre was chosen because it has been identified as the site of a major urban regeneration project. Therefore the area represents a prime candidate for predictive modelling in order to estimate what the effects of the regeneration scheme will be on crime. Prior research has already simulated the effects of the urban regeneration and was able to show that the *BurgdSim* model has utility in predicting burglary patterns at the local (household) level (Malleson, 2010). However, the previous model had no information about the burglary victims because, as Section 3.2 discussed, individual-level data were not available. Hence the following experiments have two major advantages: the model is now able to take individual household occupancy and attractiveness measures into account (previously these measures were homogeneous for all houses in an area) which brings it closer in line with criminology theory; and it is now possible to analyse the *victims* of burglary to identify which households have the highest simulated burglary risk and why.

5.2. Comparing pre- and post-integration burglary patterns

The first stage in assessing how the integration of a microsimulation model has influenced the agent-based model is to explore the change in aggregate burglary patterns. Using the expanding cell algorithm, Fig. 3 plots two error distributions. The first is the original pre-integration model errors which were calculated by comparing the simulated burglary rates to those found in real data. The second shows the difference in burglary patterns between the two (pre- and post-integration) models. Original model errors range from −2.1% to +0.8% per 0.19 km^2 cell whereas the difference between the pre- and post-integration models has a considerably smaller range between −0.1% and +0.3%. Therefore, although there are differences in the burglary patterns produced by the two models (as would be expected) these are insubstantial at an aggregate level. There are differences, however, which suggests that agents

Table 3
Results of a linear regression of house size against other variables that might increase household burglary risk. Data are for the population of victims, not the population of the simulation area. R^2 = 0.3922.

	Estimate	Std. error	*t* value	Pr (>|t|)
(Intercept)	1.7028	0.0322	52.91	0.0000
Attractiveness	−0.7696	0.0175	−44.02	0.0000
Students	1.1052	0.0131	84.57	0.0000
Family	2.0085	0.0216	93.15	0.0000
Unemployed	1.4078	0.0329	42.81	0.0000
Accessibility	0.1177	0.0225	5.23	0.0000
Visibility	0.1089	0.0517	2.10	0.0353
Traffic volume	−0.1426	0.0393	−3.62	0.0003

Please cite this article in press as: Malleson, N., & Birkin, M. Analysis of crime patterns through the integration of an agent-based model and a population microsimulation. *Computers, Environment and Urban Systems* (2012), http://dx.doi.org/10.1016/j.compenvurbsys.2012.04.003

ARTICLE IN PRESS

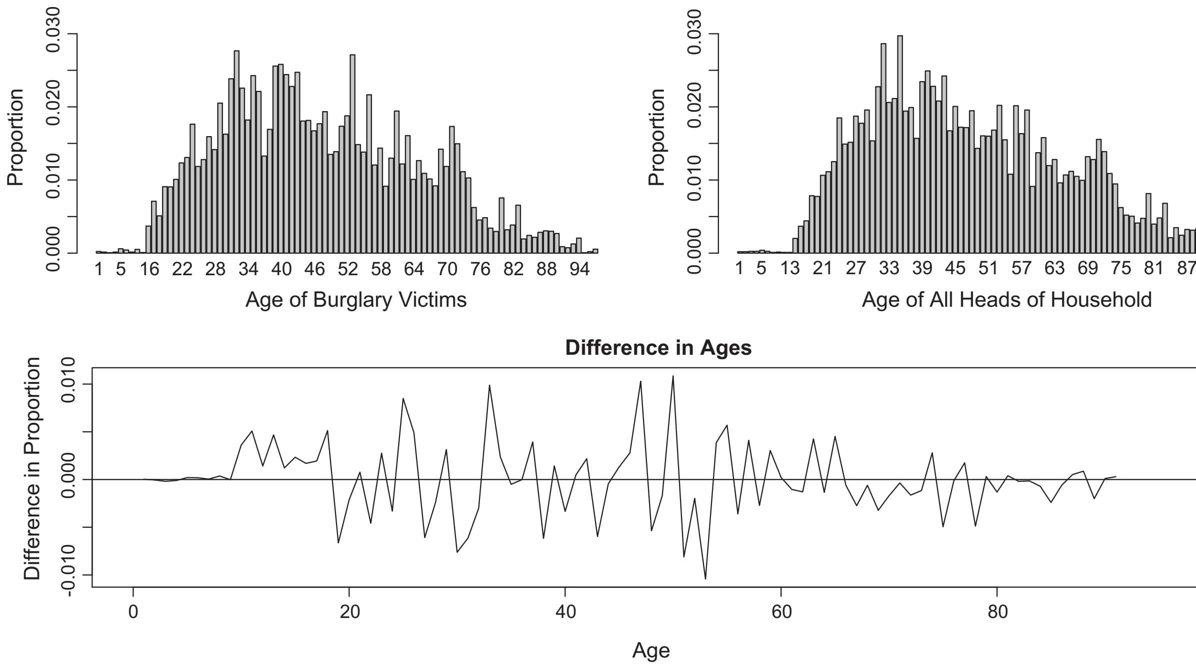

Fig. 5. Age differences between the population of victims and all households in the simulation area (heads of households only).

are choosing alternative houses or neighbourhoods, although they are not travelling to entirely different parts of the city.

5.3. Analysis of the synthetic victims

As it has been shown that incorporating individual-level victims has a relatively small affect on overall burglary patterns, a lot of information can be gained by examining the burglary victims in more detail. This was not possible before the integration of a synthetic population generated by a microsimulation model. This section will compare the attributes of the individuals who became victims of burglary in the model to the entire population of synthetic individuals. It will begin by examining the properties of the houses themselves before moving on to examine the attributes of the synthetic people.

5.3.1. Household characteristics

Fig. 4 compares the size (number of residents) and type (detached (1), semi-detached (2), terrace (3) or flat (4)) of all households in the simulation area to the subset of those that were victims of burglary (note that houses that were repeatedly victimised are represented multiple times). From the figure it becomes apparent that, in terms of house type, victims were chosen uniformly; there is no house type that has been burgled a substantially higher/lower number of times than would be expected from their proportions in the whole population. However, in terms of house size, it appears that single-occupancy houses have a higher proportion of burglaries than would be expected. This is, in itself, an interesting finding because burglar agents in the model do not take account of the number of people present in a house when making their burglary decision. It has been shown recently

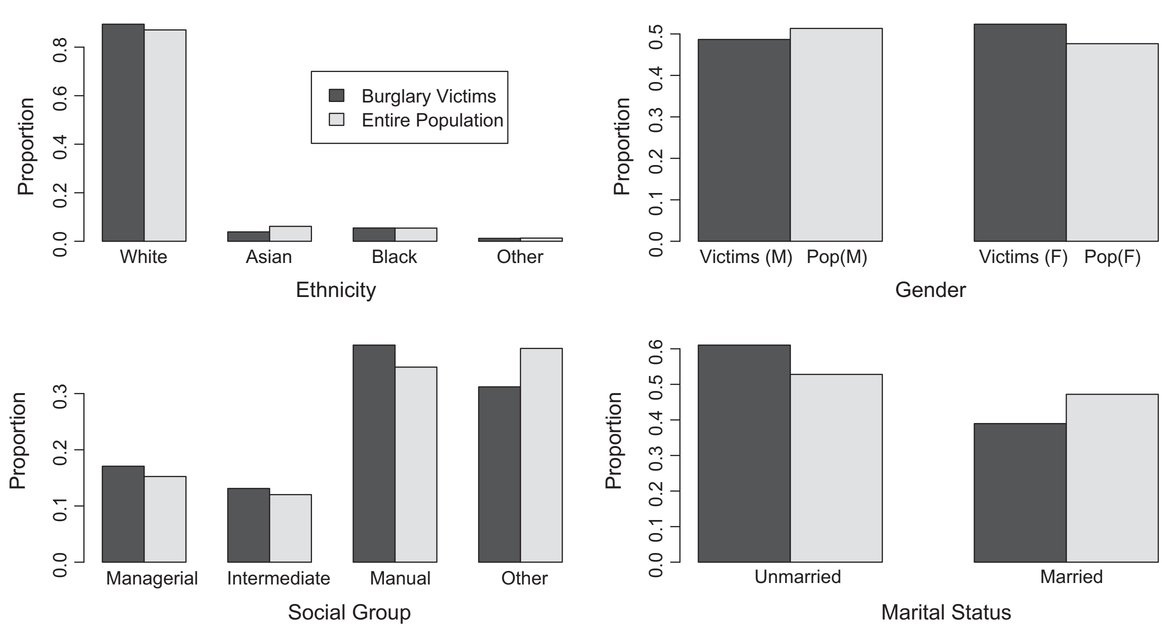

Fig. 6. Individual characteristics of the head of households for the set of victims compared to all synthetic individuals in the study area.

Please cite this article in press as: Malleson, N., & Birkin, M. Analysis of crime patterns through the integration of an agent-based model and a population microsimulation. *Computers, Environment and Urban Systems* (2012), http://dx.doi.org/10.1016/j.compenvurbsys.2012.04.003

that households with a single adult *and children* have higher burglary risks (Flatley et al., 2010), but in the synthetic data these types of family unit will be represented as a two or more person household. It is worth noting that the number of people in a household will influence the probability of the house being occupied, but this relationship is not linear – if an unemployed person or student is the sole occupant of a dwelling then their occupancy probability *increases* compared to that of a typical daytime worker.

Therefore it is likely that there is a different variable that is correlated to single-occupancy dwellings which causes households to be more vulnerable. Table 3 provides the results of a linear regression model that compares house size to the other household variables that influence burglary risk: *attractiveness* (a proxy for the social class of the residents), *occupancy* (whether the house is occupied by students, families or unemployed people), *accessibility* (how easy it is to gain access to the house), *visibility* (how visible the house is to the outside) and *traffic volume* (the estimated volume of traffic on the adjacent road). The model demonstrates that there is no clear relationship between house type and the other household variables that might influence burglary risk.

As there is no clear explanation, in terms of model rules, for the higher proportion of single-occupancy dwellings that have been burgled, it is likely that this finding is a result of the spatial configuration of the simulated area. A risk that has not been measured thus far relates to how close a house is to a potential burglar agent and whether or not the agent is aware of the house in the first place. It is likely, therefore, that houses with single occupants happen to be in areas that are in the awareness spaces of many offenders. An advantage with the use of simulation is that the researcher has full knowledge of the system that they are experimenting with and, therefore, it is hypothetically possible to record how many times a particular building falls within the awareness space of

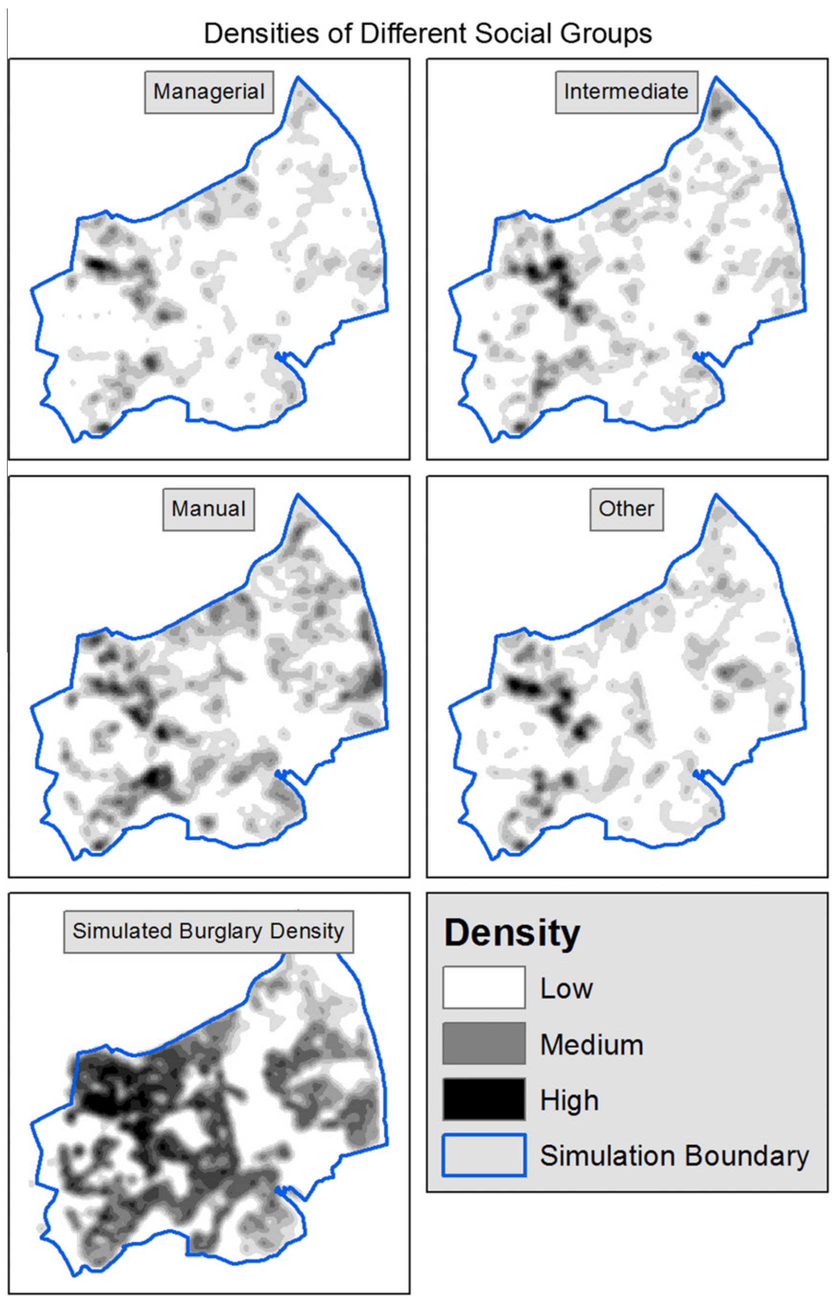

Fig. 7. Comparing the densities of different household socio-economic statuses.

the population of agents. This avenue of analysis is recommended for future work. In the meantime, the following section will explore some of the attributes of the *individual people* who have been victimised, rather than the households themselves.

5.3.2. Characteristics of individual people

To begin with, Fig. 5 illustrates the distribution in people's ages. It appears that there is relatively little difference in the age profile of the subset of burglary victims compared to that of all people in the study area. This is confirmed by the graph of the age difference; although there are differences in the proportions ages they appear to be random (there is no noticeable trend). The simulation area contains a large number of young households and also spikes in some of the more mature age categories so on the whole there is a fairly broad mix of age groups. This might be quite different to the distribution of victims found in areas that are predominantly occupied by students or the more elderly affluent suburbs.

As well as age, synthetic individuals have characteristics to represent their ethnicity, gender, social group and marital status. Fig. 6 compares these attributes for the victims and the entire population. In terms of ethnicity and gender there is little difference between the distributions of victims and non-victims. However, there are interesting variations in social group and marital status.

In terms of social group, it appears that 'manual' workers receive the highest proportion of burglaries – higher than would be expected if victims were distributed uniformly. This is surprising because burglar agents consider 'managerial' and 'intermediate' households to be more attractive than 'manual' ones. Therefore, as with the proportion of single occupancy dwellings (discussed above), it appears that the *location* of 'manual' households adds more to their burglary risk than their low attractiveness takes away. An explanation for this finding, therefore, can be sought by mapping the locations of the different social groups as in Fig. 7. Although the distributions are not completely dissimilar at an aggregate level, there are some areas where the density of 'manual' households is distinct from other groups. However, these areas do not clearly correspond to a high or low burglary rate so it is difficult to draw any firm conclusions by observing the spatial distribution in this manner.

The situation with respect to marital status is similar to that of the social groups; it appears unmarried people are targeted more often in the simulation than would be expected. This is interesting because, unlike social group, marital status plays no part in the model rules, it is purely an artefact of where (un) married people live and the types of houses/areas that they live in. It is possible that part of this relationship is related to occupancy – it is more likely that a married couple will be part of a family and, hence, have greater occupancy – but, further spatial analysis measures will need to be taken to explore this more fully.

6. Conclusions

This research has utilised two advanced computational techniques – agent-based modelling and microsimulation – in order to make progress towards an integrated micro-level crime simulation. At this stage in the research the results are promising. Environmental criminology and 'crime at places' research are highlighting the importance of analysing crime at extremely fine spatial scales (up to individual streets or households) and focussing more heavily on the behaviour of the victims of crime rather than purely the offenders. By integrating an agent-based model, that included an advanced offender behavioural framework, with detailed information about the potential victims of crime, this research has been able to produce a model that is more closely aligned with modern criminology thinking. In particular, the research was able to highlight some of the sociodemographic characteristics of the simulated victims of burglary at the household level which, before integration of the two modelling techniques, was not possible at such fine geographical scales.

A priority for ongoing work is to include greater detail in order to enhance the accuracy of the integration process. Another significant opportunity for research in the immediate future is an extended analysis of the results to look for patterns of repeat victimisation and the characteristics of the victims who are being repeatedly victimised. Research suggests that prior victimisation is the strongest determinant of future burglary victimisation – more so than any known social or demographic factors – and hence a comparison of simulated repeat victimisation to known victimisation rates will be illuminating. In particular, an analysis of repeat victimisation might shed light on the reasons for some of the phenomena that the research has not been able to fully explain, such as the tendency for houses with 'manual' occupation to exhibit a greater proportion of burglary even though model rules mean that their risk is reduced. The potential for this type of analysis is another advantage of the linking process that was not possible previously.

References

Andresen, M. A. (2009). The place of environmental criminology within criminological thought. In M. A. Andresen, P. J. Brantingham, & J. B. Kinney (Eds.), *Classics in environmental criminology* (pp. 5–28). Taylor & Francis.

Axelrod, R. (1997). Advancing the art of simulation in the social sciences. In R. Conte, R. Hegselmann, & P. Terna (Eds.), *Simulating social phenomena* (pp. 21–40). Berlin: Springer-Verlag.

Bernasco, W. (2004). How do residential burglars select target areas?: A new approach to the analysis of criminal location choice. *British Journal of Criminology, 45*(3), 296–315.

Birkin, M., Turner, A., & Wu, B. (2006). A synthetic demographic model of the UK population: Progress and problems. In *2nd International conference on e-social science*. Manchester.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences, 99003*(90), 7280–7287.

Brantingham, P. J., & Brantingham, P. (1981). Notes on the geometry of crime. In Brantingham, P. J., & Brantingham, P., (Eds.), *Environmental criminology*, (pp. 27–54). Prospect Heights.

Brantingham, P. L., & Brantingham, P. J. (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology, 13*(1), 3–28.

Castle, C. J. E., & Crooks, A. T. (2006). *Principles and concepts of agent-based modelling for developing geospatial simulations*. <http://eprints.ucl.ac.uk/archive/00003342/01/3342.pdf> (UCL Working Papers Series, Paper 110, Centre For Advanced Spatial Analysis, University College London).

Clarke, R. V., & Cornish, D. B. (1985). Modeling offenders' decisions: A framework for research and policy. *Crime and Justice, 6*, 147–185.

Cohen, L., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review, 44*, 588–608.

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

Cromwell, P. F., Olson, J. N., & Avary, D. W. (1991). *Breaking and entering: An ethnographic analysis of burglary*. *Studies in Crime, Law and Justice* (Vol. 8). Newbury Park, London: Sage Publications.

Eck, J. E., & Liu, L. (2008). Contrasting simulated and empirical experiments in crime prevention. *Journal of Experimental Criminology, 4*(3), 195–213.

Eck, J., & Weisburd, D. (1995). *Crime places in crime theory*. *Crime and place* (Vol. 4). Newbury Park, London: Sage Publications.

Epstein, J., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. Newbury Park, London: Criminal Justice Press.

Flatley, J., Kershaw, C., Smith, K., Chaplin, R., & Moon, D. (2010). *Crime in England and Wales 2009/10. Findings from the British crime survey and police recorded crime* (3rd ed.,). London: Home Office.

Glyde, J. (1856). Localities of crime in suffolk. *Journal of the Statistical Society of London, 19*, 102–106.

Harland, K., Heppenstall, A., Smith, D., & Birkin, M. (2012). Creating realistic synthetic populations at varying spatial scales: A comparative critique of microsimulation techniques. *Journal of Artificial Societies and Social Simulation, 15*(1).

Heppenstall, A. J., Evans, A., & Birkin, M. H. (2007). Genetic algorithm optimisation of an agent-based model for simulating a retail market. *Environment and Planning B: Planning and Design, 34*, 1051–1070.

Jeffery, C. R. (1971). *Crime prevention through environmental design*. Sage Publications.

Please cite this article in press as: Malleson, N., & Birkin, M. Analysis of crime patterns through the integration of an agent-based model and a population microsimulation. *Computers, Environment and Urban Systems* (2012), http://dx.doi.org/10.1016/j.compenvurbsys.2012.04.003

Johnson, S. D., & Bowers, K. J. (2009). Permeability and burglary risk: Are Cul-de-Sacs Safer? *Journal of Quantitative Criminology, 26*(1), 89–111.

Kongmuang, C. (2006). *Modelling crime: A spatial microsimulation approach*. PhD thesis, School of Geography, University of Leeds, LS2 9JT, UK.

Malleson, N. (2010). *Agent-based modelling of burglary*. PhD thesis, School of Geography, University of Leeds, UK.

Malleson, N., Heppenstall, A., Evans, A., See, L. (2010). *Evaluating an agent-based model of burglary*. Working paper 10/1, School of Geography, University of Leeds, UK. <http://www.geog.leeds.ac.uk/fileadmin/downloads/school/research/wpapers/10_1.pdf>, January 2010.

Malleson, N., Heppenstall, A., & See, L. (2010). Crime reduction through simulation: An agent-based model of burglary. *Computers, Environment and Urban Systems, 34*(3), 236–250.

Malleson, N., See, L., Evans, A., & Heppenstall, A. (2012). Implementing comprehensive offender behaviour in a realistic agent-based model of burglary. *Simulation: Transactions of the Society for Modeling and Simulation International, 88*(1), 50–71.

Openshaw, S. (1984). *The modifiable areal unit problem*. *Concepts and techniques in modern geography*. Norwich: Geo Books.

O'Sullivan, D., & Haklay, M. (2000). Agent-based models and individualism: is the world agent-based? *Environment and Planning A, 32*(8), 1409–1425.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*, 918–924.

Schelling, T. (1969). Models of segregation. *The American Economic Review, 59*(2), 488–493.

Schmidt, B. (2000). *The modelling of human behaviour*. Ghent, Belgium: Society for Computer Simulation International.

Tseloni, A. (2006). Multilevel modelling of the number of property crimes: Household and area effects. *Journal of the Royal Statistical Society A, 169*, 205–233.

Wiles, P., & Costello, A. (2000). *The 'road to nowhere': The evidence for travelling criminals*. *Home Office Research Study 207*. London: Home Office.

Williamson, P., Birkin, M., & Rees, P. H. (1998). The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning A, 30*, 785–816.

Wright, R., & Decker, S. (1996). *Burglars on the job: Streetlife and residential break-ins*. Boston: Northeastern University Press.

Wu, B., Birkin, M., & Rees, P. (2008). A spatial microsimulation model with student agents. *Computers, Environment and Urban Systems, 32*(6), 440–453.

Wu, B. M., Birkin, M. H., & Rees, P. H. (2010). A dynamic MSM with agent elements for spatial demographic forecasting. *Social Science Computer Review, 29*(1), 145–160.

Please cite this article in press as: Malleson, N., & Birkin, M. Analysis of crime patterns through the integration of an agent-based model and a population microsimulation. *Computers, Environment and Urban Systems* (2012), http://dx.doi.org/10.1016/j.compenvurbsys.2012.04.003