The Development of An Urban Crime Simulator

*Final Report*

Jay Lee  
*Urban Operations Research, Inc.*  
671 River Bend Blvd  
Kent, OH 44240

1\textsuperscript{st} version: November 18, 2009  
2\textsuperscript{nd} version: December 1, 2009
Executive Summary

The development of an Urban Crime Simulator has resulted in a software tool that allows users to simulate changes in crime rates of urban neighborhoods with his/her own GIS dataset, to select data attributes from the GIS dataset, to assign weights to the attributes based on his/her professional experience or knowledge acquired from literature. Urban Crime Simulator was developed using C#, a computer programming language as a part of Microsoft Visual Studio, and library routines in ArcEngine. Urban Crime Simulator groups urban neighborhoods based on users’ definition of neighborhood characteristics into clusters and uses the resulting clusters as the basis for estimating changes in crime rates. Users update individual urban neighborhoods with projected/simulated growth such for Urban Crime Simulator to estimate changes in crime by searching through clustered neighborhood for the neighborhood with the most similar profile. The crime rate of the neighborhood with the most similar profile as the updated neighborhood suggests an expected level of crime as the result of simulation.

Through this developmental effort, we argue that a combination of criminological theory, coupled with the concept of neighborhood life cycles, are a better approach to estimating/simulating changes in crime rates in urban neighborhoods. We recognize the limitation of using commercially available routine libraries as it requires royalty fees when distributing the resulting software tools. However, given the short time and limited budget, other options are limited. Finally, with the flexible modeling structure and flexible level of geographic aggregation that UCS accepts GIS data, we suggest that additional data mining tools be added to assist users of UCS better understand spatial and temporal trends of their urban neighborhoods before and during the simulation of urban crime.

Literature Review

The development for an Urban Crime Simulator (Simulator, or UCS) was originated from the idea that simulated urban growth may be used to estimate changes in crime rates of neighborhoods. There have been many attempts for simulating urban growth but most have used the raster data structure instead of the vector data structure that corresponds to the format of data that most law enforcement agencies use. An exception of this trend is the Urban Growth Simulator (UGS) developed at the Kent State University Applied Geography Laboratory with funding provided by the US Environmental Protection Agency. Urban Growth Simulator (UGS) has two versions: an online version that is tied to a database of the 15-county region in Northeast Ohio (http://empact.geog.kent.edu) and a stand-alone version (available in the /UCS/Software/Urban Growth Simulator 2.4/ folder, or downloadable from
An UGS workbook is also included as a reference for how-to’s and as a tutorial.

UGS works with vector GIS data in ways that allow users to specify estimated magnitude of growth with a number of growth management tools to simulate how the estimated growth will be distributed geographically. In the simulation process, UGS uses a series of procedures for identifying developable land, complying with local zoning codes, avoiding environmentally critical areas, and incorporating neighborhood plans. The output from UGS can be stored as shapefile, a GIS data format by ESRI, Inc. (Redlands, California). The shapefile that has the simulated growth areas can then be overlaid with other GIS data layers to identify neighborhoods in urban areas that may be seeing growth in the time period simulated.

With urban neighborhoods identified to receive future growth, our plan was to develop an urban crime simulator (UCS) for estimating how crime rates would change if neighborhoods were estimated to have growth. As the first attempt, we defined that property crime was to be the focus and that census blockgroup was to be the unit of analysis. Property crimes, unlike other types of crime, seem to have more geography in the way they are located and distributed. Census blockgroup, as an analytical unit, was selected because it offers more details than the widely used Census tract level and, at the same time, have more attribute data available than Census blocks.

A literature review was carried out in Phase One of this project. Full text of this literature review is included on this DVD (/UCS/Documents/Reports/Literature Review.PDF). Given the purpose for developing an Urban Crime Simulator, it is necessary to review what criminological theories can provide and support for the estimation of changes in crime rates as induced by urban growth. At the same time, it is also necessary to review what simulation approaches are available and are being used with what level of success as well as how suitable they are for the issue at hand.

A summary of findings from the literature review on this topic is provided below:

- For a theoretical framework of simulating property crime, we found that:
  - No single theory works for such complex phenomena. A combination of the Routine Activity theory, Social Disorganization theory, and Deviant Place theory works well for the purpose. These three theories provide the conceptual basis for addressing various components needed to explain the sophisticated interplays between offenders, properties, and environments.
  - Recognizing that not all neighborhoods possess the same environmental attributes and that not all neighborhoods change in the same pace or in the same direction,
we incorporated the concept of Neighborhood Life Cycle that suggests neighborhoods evolved through a set of developmental stages. This concept provides the needed framework for our simulation algorithm. It works perfectly with the three criminological theories mentioned above.

In terms of the simulation approaches, we have reviewed cellular automation, agent-based modeling, multi-agent based simulation that seem to have captured researchers in recent studies in criminology, GIS, and related fields. Overall, the review suggests that the raster data structure that is used by most of these simulation approaches is not suitable for the development of UCS because most users of UCS would be using vector GIS data. While it is possible to convert data from vector data format to raster format, it should be noted that such conversion typically introduces unnecessary distortion of data in terms of precision and accuracy.

A summary of findings from the literature review on this topic is provided below:

- Cellular Automation (CA) requires that the study area be partitioned into a raster grid of regularly sized cells. Each cell is assigned to have one of the states. With a set of rules governing how each state changes to other states, cells in the grid change their states based on the set of rules and how analysts structured the rules. It is not a feasible approach for UCS because of the limitation that CA works in raster data structure. In addition, the temporal dimension of the rule set is defined by the analyst, which in turn, is subject to analyst’s own expertise and to particular locality, making it not possible to generalize results from one area to others.

- Agent-based modeling (ABM) and multi-agent based simulation (MABS) have received increasing attention in recent years as they are applied or tested in different fields of research. So far as what existing studies suggest, the level of details is necessarily high for the set of rules that govern how individual agents behave in the system. At the same time, both simulation approaches need temporal rules for agents to know when to initiate or terminate their assigned actions. Current literature in criminology provides insufficient knowledge at the level of details needed for such rule set, making the adoption of ABM or MABS not practical. It is also not feasible to assume that users of UCS will have time to define the temporal aspects of all individual agents in the simulation so they will function.

- With what we learned from CA, ABM, and MABS, it seems necessary to develop a new algorithm for use in the Urban Crime Simulator.
Aggregating individual events to various geographic units has been a topic for research in many fields. In Geography, modifiable area unit problems show that relationships between variables may be weakened if individual data are aggregated from smaller geographic units to larger ones. In other fields, this problem is known as ecological fallacy. In considering the development of the UCS, inevitably the issue of geographical aggregation of data has to be addressed.

A summary of findings from the literature review on this topic is provided below:

- Geographic aggregation tends to generalize the variation of whatever phenomena being studies within each geographic unit. It is apparent that as smaller geographic units are used in a study, the more details can be revealed in terms of how the phenomena vary over space. However, the availability of data at smaller geographic units is often the biggest barrier that researchers or analysts face. The availability of the required data often is a major force that dictates many studies utilizing data that are collected by government agencies.

**Methodology**

Given results of the literature review, we have developed a new methodology for UCS. The methodology is further explained in the Methodology.PDF, also included in this DVD within the /UCS/Documents/Reports/ folder).

A summary for the issued considered when developing the methodology is provided below:

- Simulation process should be flexible to allow users to use various modeling approaches
  - Given the wide spectrum of criminological theories and the varying success of each theory in explaining the geography of crime, it is necessary to formulate UCS to be flexible such that users can incorporate their own modeling structure, expertise, or professional experience.

- Data assembly should not be a barrier for using the Simulator
  - Data availability dictates the feasibility in many analytical procedures and simulating urban crime is no exception. UCS should be designed such that it will allow users to use it with whatever data they may have, if not all possibilities, at least those using data formulated for census blocks, blockgroups, tracts, counties, or other polygonal units.

- GUI should be as simple as possible to shorten learning curves
  - Using newest software development tools, UCS should be designed to be as intuitive to use as possible while avoiding misuses. With most law enforcement officers being busy
with other responsibilities, it is desirable that UCS be a simple-to-use tool that does not demand spending long time to learn to use.

- Simulation algorithm should be robust and flexible
  - UCS should be developed with a simulation algorithm that is robust with the same outcome when given the same conditions. Furthermore, UCS should provide tools to users to allow different attributes and weights in their simulation models.

In a summary form, the simulation methodology is a four-step process:

1. Analysts (or UCS users) define the neighborhoods, attributes of neighborhoods, and the weights of neighborhoods.
2. Based on the included neighborhood attributes and their weights, all neighborhoods are grouped into a number of clusters, using multivariate cluster analysis. Homogeneity within each cluster is maximized while heterogeneity between clusters is maximized.
3. Once neighborhoods are classified into clusters, individual neighborhoods can be selected and updated with estimated changes from simulated urban growth in these neighborhoods.
4. One at a time after each neighborhood is updated with simulated urban growth, UCS searches through all neighborhoods in the studied area to find one neighborhood that has the most similar profile with the updated neighborhood, based on the included attributes and weights. The crime rate, or level of crime, of the neighborhood found to be the most similar to the updated neighborhood is outputted as the estimated crime rate after the neighborhood receives the simulated urban growth.

**Development and Assessment**

Software development for UCS went through three stages as outlined below:

- Initially, FORTRAN was used to develop library routines for statistical analysis (multivariate cluster analysis). The library routines were tested and converted to software components that were incorporated into the Simulator.
- ArcGIS VBA was used for the first round of software development. While successful, it was discarded as there was no way to block user access to source code. Blocking user access to source code is mainly for prevention of accidental changes to how Simulator functions that may resulted in undesirable outcomes that users may not be aware of. Access to source code can be granted by the funding agency or with simple requests made to us.
We eventually chose to use Microsoft Visual Studio and ArcEngine as the platforms for the development of the Simulator. Developed UCS was distributed to students in GIS classes at Kent State University. Without any instruction or training (but with a prepared GIS polygon shapefile), students were able to select neighborhood attributes, assign weights, classify neighborhoods, and simulate changes in crime rates for neighborhoods with updated attribute information. UCS was presented and discussed at the 2009 Applied Geography Conference held in Baton Rouge, Louisiana from 10/28 to 20/30. A total of four conference participants tried UCS at the conference without any problem.

Our assessment of the software development is that the programming aspect would need more time to fine tune the code and that there may be additional functions UCS can be added to help users explore their data to better understand their neighborhoods. The current version of UCS assumes that users possess a sufficient level of understanding/familiarity of the neighborhoods they work with so that they can meaningfully select neighborhood attributes and assign their weights for the simulation model. While this is a reasonable assumption, UCS could be equipped with additional data mining tools to assist users in exploring spatial trends or temporal variation of crimes in their urban areas before simulating changes.

Specific to programming UCS, we chose to use Microsoft Visual Studio for maximizing future compatibility as Microsoft Visual Studio seems to be the most dominating programming tool on PC platform. In terms of our selection of using ArcEngine routine libraries, we chose to use ArcEngine because of the readily available software components and their compatibility with the most widely used GIS data format. Given the short time and limited budget available for this project, using ArcEngine allows us to avoid spending much time and effort to re-program the software components that ArcEngine provides.

**Concluding Remarks**

Although several simulation approaches, such as CA, ABM, MABS, among others, have been demonstrated to be useful tools in simulating various geographic phenomena, it is clear from the literature review conducted for this study that they are limited to specific data structure and specific conditions. The lack of possible generalization from CA, ABM, MABS, and other simulation approaches prompted us to a different direction which resulted in the simulation algorithm we used for UCS.
The criminological literature suggests a wide spectrum of theories that may be useful in estimating changes in crime rates in urban neighborhoods. However, it seems not one theory is comprehensive in dealing with all aspects of this process.

From this project we have the following experience and thoughts for future development:

- That a synthesis of multiple theories works better for complex phenomena such as urban crime and its geographic distribution than any single theory does.
- That software tools such as UCS need to have the flexibility of allowing users to use data that are available to them, which may be at various levels of geographical aggregation.
- That local knowledge is the key to successful and meaningful simulation of urban crime.
- That a better suite of data mining tools may be added to UCS to help users explore spatial and temporal trends of their GIS data before defining their simulation models in UCS.