C

Calibration of Simulation Models

Nick Malleson
School of Geography, University of Leeds, Leeds, UK

Overview

Simulation models are becoming increasingly popular in criminology research. In order for researchers to have confidence in the results of simulation studies, it is essential to make sure that the models are properly evaluated. *Calibration* is a major element to this evaluation and refers to the estimation and adjustment of model parameters to improve the agreement between model output and a data set. This entry will discuss some different methods of model calibration as well as related quantitative methods of error assessment.

Introduction

Simulation models are becoming increasingly popular in criminology research, both as tools for exploring real-world crime patterns (explanatory models) and for experimenting with underlying theory (conceptual models). In order for researchers to have confidence in the results of simulation studies, it is essential to make sure that the models are properly evaluated. The process of evaluation is commonly divided into three activities, defined by Rykiel (1996) as follows:

- **Verification** is a demonstration that the modeling formalism is correct.
- **Calibration** is the estimation and adjustment of model parameters and constants to improve the agreement between model output and a data set.
- **Validation** is a demonstration that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model.

It is possible to calibrate a model quantitatively by assessing error using statistics or qualitatively by manually comparing model results and field data. Qualitative approaches are particularly well suited to spatial models where the researcher is able to compare maps. However, for complicated models with many parameters, a qualitative approach to calibration will probably be very time consuming and unlikely to reveal the optimal model configuration (which is the ultimate aim of calibration). Therefore, this article will focus on quantitative methods of error assessment (rather than relying on human objectivity) and automatic calibration routines that are able to explore a model's parameter space and estimate error without human intervention. Also, in the context of environmental criminology, it is much more important to accurately reflect field conditions when working with *explanatory models* – i.e., those that simulate real-world conditions – so the discussion will focus on these in particular. Conceptual models,

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

on the other hand, will not necessarily attempt to replicate real environmental conditions so automatic calibration methods are less relevant.

The article is organized as follows. Section outlines the main principles of calibration and Section follows with a more in-depth assessment of a number of methods for model calibration. Sections and complete the article by discussing the state of the art, literature controversies, and open questions.

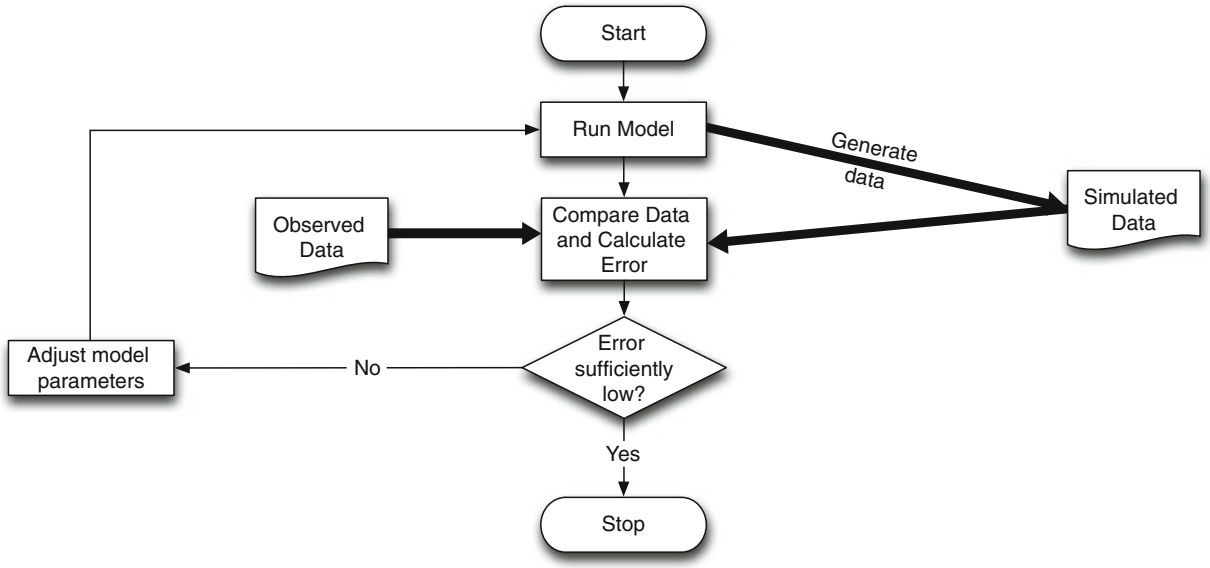
Background

Calibration refers to the process of configuring a model's parameters to match some observed historical data. This usually consists of searching for a combination of parameter values that cause the model to produce data which are similar to that collected from the real system under investigation. In many cases, a single "fitness" value is sought which succinctly summarizes the correspondence between simulated data and field observations.

Figure 1 illustrates the calibration process. The model is repeatedly reconfigured with the aim of reducing the error between the results and the field data. Once a predetermined error level has been reached, the process ends and the model can be considered calibrated, i.e., configured in such a way that it is apparently able to simulate the real system effectively. The level of error deemed acceptable is subjective and depends on the individual study. It is also worth noting that simply matching data is not necessarily a sufficient criteria for establishing model correctness, which is a point Section will address in more detail.

With simple models, the process of configuring parameters and calculating error is usually relatively simple. However, *simulation* models are often extremely complex and can contain a large number of configurable parameters. To confound the situation, model parameters often have nonlinear effects on the model's behavior which makes it difficult to predict how the model will behave under new parameter configurations. Hence, the process of manipulating a model's parameters to match some field conditions is often nontrivial.

To being with, the means of evaluating the degree of similarity between model results and expected data can be troublesome in itself. Once a quantitative measure of similarity has been developed, there are numerous methods that can be used to explore the parameter space of the model in search of the optimal configuration. These are also known as *optimization* methods. The following section will first discuss the



Calibration of Simulation Models, Fig. 1 An overview of the process of calibration

different methods that can be used to calculate error and follow with a review of a number of automatic calibration/optimization routines.

Methods for Model Calibration

Assessing Goodness of Fit

In the context of calibration, goodness of fit (GoF) is an important measure. It refers to the error between the results of a model and the data that it is trying to replicate (hereafter referred to as "observed data"). Commonly, GoF statistics are applied to tabular data, such as the number (or rate) of crimes against different types of people, within different spatial areas or during a particular time period. Knudsen and Fotheringham (1986) experimented with a number of goodness-of-fit statistics and found the standardized root mean square error (SRMSE) to be the best performing. A drawback with SRMSE, however, is that the value itself is difficult to understand. For example, with the SRMSE it is not possible to state what percentage of the variation in observed data can be accounted for by a model. An alternative statistic, R^2 solves this problem because it represents the percentage of agreement between the model and the expected data. However, R^2 is insensitive to the overall amount of error, predicting a good fit in some circumstances where the SRMSE would not (Harland 2008).

The SRMSE can be defined as

$$SRMSE = \frac{\sqrt{\left(\Sigma\left(y_i' - y_i\right)^2/n\right)}}{\bar{y}} \qquad (1)$$

where y_i' is the predicted value at matrix point i, y_i is the actual value at i, $\bar{y}$ is the mean value of the predicted values (y'), and n is the total number of values. The lower limit of the statistic is 0 which indicates no difference between the predicted values (y_i') and the observed values (y_i). The upper limit is usually 1 (Knudsen and Fotheringham 1986) but can be greater, particularly when matrices are sparse (Harland 2008).

Using the same notation, R^2 can be defined as

$$R^2 = 1 - \frac{\sum_i \left(y_i - y_i'\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2} \qquad (2)$$

and a value of 1 indicates identical data sets. The lower limit of the statistic is 0.

There are a variety of alternative GoF statistics that could be used to assess model error and the most appropriate will depend on the data and application area. For more information about the statistics discussed here, the interested reader can refer to the comprehensive assessment in Knudsen and Fotheringham (1986). Alternatively, there are a wide range of textbooks that define methods for both parametric and non-parametric data. In terms of calibration, the most important decision is to choose the GoF statistic that is appropriate to the study and the nature of the data so that a reliable assessment of error can be made. In addition, more than one statistic could be used simultaneously to provide a more comprehensive assessment of model error.

Goodness of Fit for Spatial Data

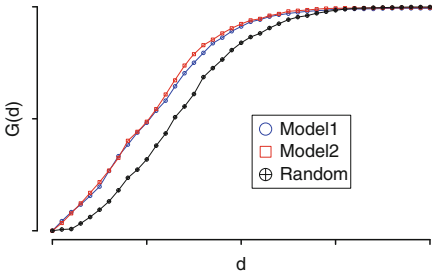
Many simulation models in environmental criminology explore the spatial distributions of crime (Liu et al. 2005; Groff 2007; Hayslett-McCall et al. 2008; Dray et al. 2008; Birks et al. 2012; Malleson et al. 2013). If the simulations work at an aggregate spatial scale (measuring crimes per area), then the procedure for assessing GoF is the same as that for nonspatial data. However, it is preferable for simulation models to use data on individual crime occurrences and, therefore, generate point pattern data. Common GoF measures cannot be used to compare point patterns directly as the data must be in the form of a table or matrix. One solution to this problem is to first aggregate the point data to commonly used areal boundaries such as the enumeration district or census tract. However, this process is far from ideal. Firstly, such aggregation will expose the results to the modifiable areal unit problem (MAUP: Openshaw 1984). Openshaw found

that changing the size and shape of the boundaries themselves can have a dramatic effect on the resulting spatial patterns and subsequent results. Secondly, the process of aggregating is likely to hide important patterns that are present at finer geographies. Andresen and Malleson (2011), for example, showed that there was considerable spatial heterogeneity in crime rates at the street level which would be hidden at larger spatial scales. Hence aggregation will inevitably introduce error.

An alternative approach is to compare the point patterns directly *without aggregating*. Although no commonly accepted methods exist for this purpose, such as the SRMSE or R^2 there are a a number of useful spatial statistics that can describe the distributions of point patterns and can be used to make mathematical comparisons. These could replace, or compliment, traditional GoF statistics used to determine model error during calibration. Table 1 summarizes some of these statistics and illustrates the results of their application to three data sets: two similar point patterns produced by a simulation model ("Model1" and "Model2") and a point pattern produced by a random process ("Random").

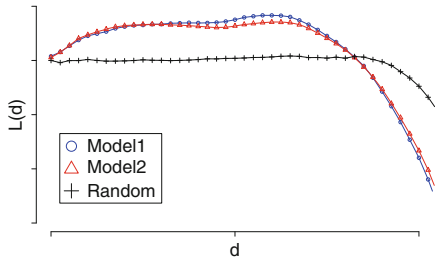
The functions outlined in Table 1 provide information about the degree of clustering in

Calibration of Simulation Models, Table 1 A summary of spatial statistics that can be used to describe and compare the spatial structure of point patterns (Malleson 2010)

Statistic	Pros/cons	Usage with example data
Nearest Neighbour Index (NNI) – also known as the Clark and Evans *R* statistic (Clark and Evans 1954) – is the ratio of the minimum nearest-neighbour distance. The nearest-neighbour distance for a point *i* is the distance to the closest neighbouring point. ($d_{\min}$) to the expected minimum distance for a random point pattern ($\bar{\delta}$) $NNI = \frac{\bar{d}_{\min}}{\bar{\delta}} = \frac{\sum_{i=1}^{n} \frac{d_{ij}}{n}}{\frac{1}{2\sqrt{A/n}}}$ (3)	Gives a concise general picture of whether or not clustering is present (compared to random data) Usefil as a preliminary procedure (Bailey and Gatrell 1995) It is difficult to account for edge effects (there are some solutions to edge effect problems, such as circular or rectangular corrections (Levine 2006), but these are not ideal (Chainey and Ratcliffe 2005)) Too simplistic to be really useful	The NNI statistics suggests that all the data used in the following examples (Modell, Model2 and Random) are clustered which is to be expected (Malleson et al. 2010). Althought the statistic is not comprehensive enough to assess error in isolation, it could be useful as a preliminary measure of similarity as part of a larger error assessment during calibration
The *G* function, at a given distance, *d*, is the fraction of points, s_i, whose nearest neighbour is less than *d* away: $G = \frac{\#(d_{\min}(s_i)<d)}{n}$ (4) where # means "the number of" (as in Bailey and Gatrell 1995) **The *F* function** is similar to *G* but uses the distance from a randomly selected map location to the nearest point. As defined bu O'Sullivan and Unwin (2003): if $\{p_1 \ldots p_i \ldots p_m\}$ is a set of *m* randomly selected locations and *S* is the set of all points, then: $F = \frac{\#(d_{\min}(p_i,S)d)}{m}$ (5)	Describe clustering in more detail than the NNI index by providing a measure at different distances Can be used to differentiate between clustered and uniform data (see O'Sullivan and Unwin (2003) for a fuller discussion) Only consider a *single* nearest neighbour distance in their calculations so disregard a considerable amount of information	The following illustrates the *G* functions for the Model1, Model2 and Random data sets (graphs of *F* are similar). The graphs illustrate that, at shorter distances, the model data sets are more clustered than the Random one. Therefore a traditional GoF statistic could subsequently be used to estimate the differences between the *G* of *F* graphs for different data. This would provide a single statistical measure of similarity which could be used during calibration 

(*continued*)

Calibration of Simulation Models, Table 1 (continued)

Statistic	Pros/cons	Usage with example data
The *L* function is a transformation of Ripley's *K* that provides evidence for whether or not clustering is more or less than would be expected under complete spatial randomness (CSR) (O'Sullivan and Unwin 2003): $L(d) = \sqrt{\frac{K(d)}{\pi}} - d$ (6) Values of $L(d) < 0$ suggest that there are fewer events in the space than would be expected under CSR and that the data are therefore less clustered. The reverse is true for $L(d) > 0$	Takes all the neighbours that are within a given distance into account so is a more descriptive statistic By comparing graphs of $L(d)$ it is possible to determine how similar the clustering of to point patterns is	The following graph illustrates that the Model1 and Model2 data are more clustered than the random dataset (which equates to approximately $L(d) = 0$ for low d values). Above $d \approx 3{,}000$ L begins to fall due to boundary effects (this is because many of the large circles produced by the underlying K function are nearly empty at large distances because they cover areas outside the simulation boundary with no points (O'Sullivan and Unwin 2003). As with the G and F functions, it would be possible to use a traditional GoF statistic to quantitatively compare the similarity of L functions for two point patterns (e.g., simulated data and calibration data) 

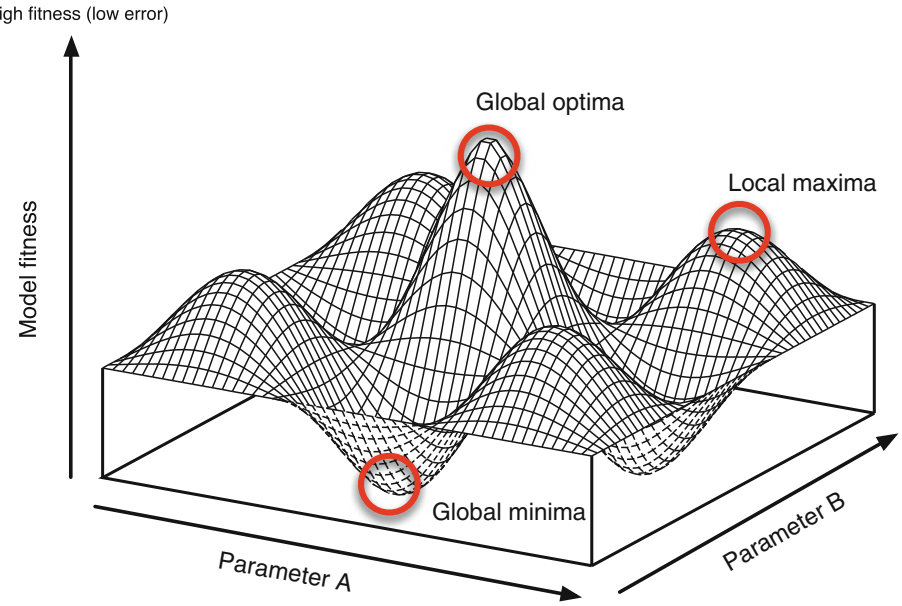
point pattern data. Although these statistics have the benefit that they are not susceptible to the modifiable areal unit problem, they could not be used during calibration, in isolation, because it is possible that two different point patterns nevertheless have the same clustering properties. However, these could be used in conjunction with other methods (such as aggregating the points to area boundaries and applying a GoF test) to provide a more comprehensive error assessment.

As well as comparing mathematical descriptions of clustering in point patterns, it is also possible to generate raster density maps from the point patterns and compare these mathematically. This approach is common in the field of spatial modeling for comparing simulated and real land use. For a review of recent approaches, the reader is directed to Kuhnert et al. (2005). Finally, it is also possible to aggregate the point patterns to a regular grid and then use traditional GoF statistics on the resulting matrix as discussed by Costanza (1989). This approach has the advantage of reducing the effects of the modifiable areal unit problem because numerous regular grids can be applied to the point patterns at the same resolution.

Exploring the Parameter Space

Having determined an effective means of estimating the goodness of fit between simulated and observed data, calibration itself can begin. The task of model calibration is effectively a search through the model parameter space, assessing the accuracy of the model under different combinations of parameters. For simulation models with large numbers of continuous parameters, this search space will be extensive. This is confounded by the fact that simulation models are often nonlinear, so the effects of changing parameter values are not easy to predict.

For example, Fig. 2 presents an example parameter space for a nonlinear model with two continuous parameters (A and B). Varying parameter combinations changes the accuracy of



Calibration of Simulation Models, Fig. 2 An example parameter space for a nonlinear model with two parameters

the model results; hence the aim of the calibration process is to find the optimal parameter configurations ("global optima"). However, even for a model with a low number of parameters, it becomes apparent that a manual search of the parameter space is unlikely to reveal global optima. Therefore calibration is often conducted by computer algorithms which are able to intelligently search the vast parameter space.

Parameter Sweeps

A parameter sweep is a simple process of systematically varying the model parameters in sequence so that many possible combinations are explored. Each parameter to be tested has a *start*, *end*, and *increment* value which determines the range of possible values that the parameter will take. Again using Fig. 2 as an example, a parameter sweep with $start = 1$, $end = 10$, and $increment = 1$ for both parameters A and B results in 100 possible parameter combinations.

Although the method is simple, it has a number of drawbacks, namely:

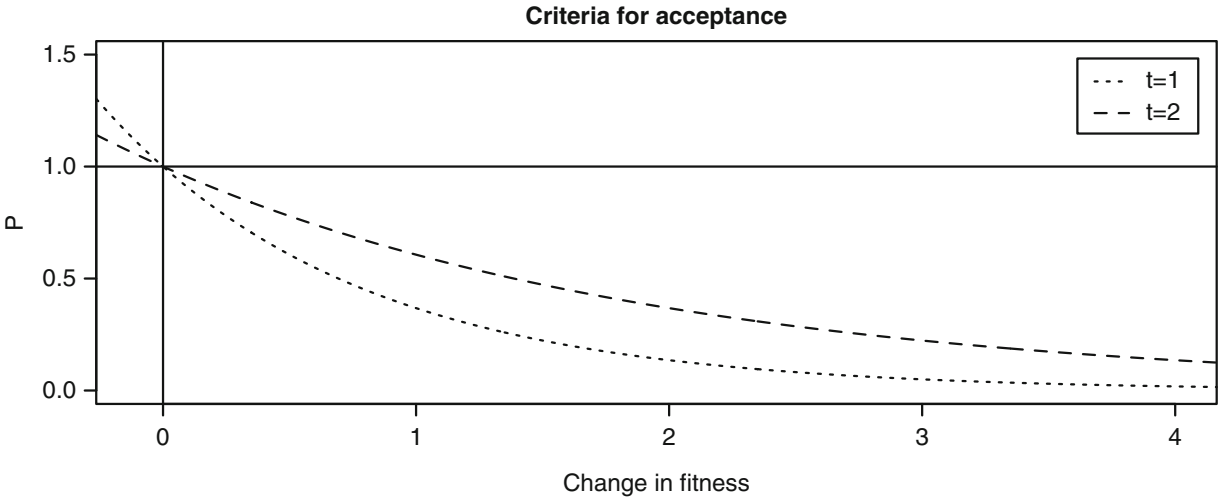
- The number of individual runs can be extensive because it increases exponentially with the number of parameters. For example, to conduct a parameter sweep on a model with five parameters, each of them integers between 1 and 10, a total of $10^5 = 100{,}000$ model runs will be required.
- The sweep explores the entire parameter space equally. More advanced approaches are able to expend greater effort fine tuning already successful configurations and ignore parts of the parameter space that are extremely unlikely to reveal global optima.

Hill Climbing

Hill climbing is a procedure that starts with an arbitrary model configuration and makes a small change to one of the parameters. If the change improves the performance of the model (reduces error), then it is retained. If the change increases error, then it is discarded. This process repeats until there are no possible parameter changes that improve the model. The main drawback with the approach is it is likely to get stuck on suboptimal configurations (e.g., the local maxima in Fig. 2), in which case it would be necessary to temporarily accept changes to the configuration that will actually *increase* the error between the model and the calibration data.

Simulated Annealing

Annealing is the process used in metallurgy in which a material is heated and then cooled in order to alter its properties (strength, brittleness, etc.). The amount of time spent at a high temperature and the rate of cooling influence the resulting properties of the material. The



Calibration of Simulation Models, Fig. 3 The effect of temperature on the probability of accepting a regressive move

simulated annealing optimization procedure takes its inspiration from an algorithm that simulates this annealing process (Metropolis et al. 1953). Originally formulated by Kirkpatrick et al. (1983), the procedure improves upon hill climbing algorithms by occasionally allowing choices that *lower* the fitness of the model. Whereas hill climbing will always choose the best move from those available, simulated annealing chooses a random move from the neighborhood. If the move improves the fitness, then it is always accepted, but if it does not, then there is still a *possibility* that it will be accepted. This helps the algorithm to climb out of local maxima.

The criteria for accepting a lower fitness is given by

$$P = e^{-c/t} > R(0,1) \tag{7}$$

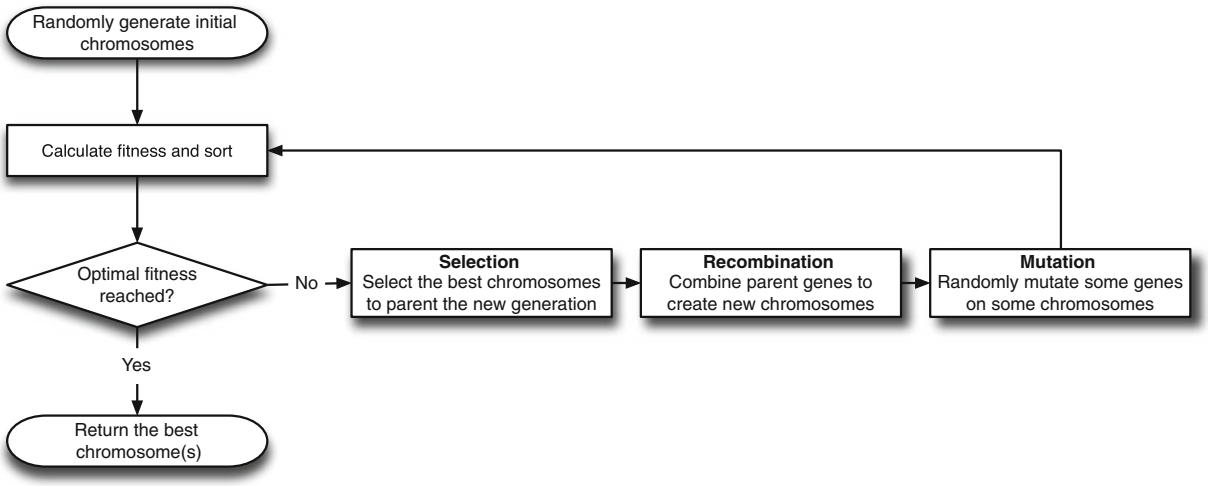
where c is the change in fitness (negative for an improvement, positive for deterioration), t is the temperature, and $R(0,1)$ is a random number in the range 0–1. The temperature is used to reduce the chance of accepting poor moves over time so that the algorithm will converge. This is analogous to gradually reducing the temperature in the annealing process. If $t = 0$ then only improvements to the fitness will be accepted which causes the algorithm to behave like a hill climbing procedure. Figure 3 illustrates the change in P for a range of moves under two different temperatures. Note that if the change is positive ($x < 0$), then $P > 1$ so the move will always be accepted. As t decreases, so does the probability of accepting a move.

The rate at which t drops, as well as its initial and final values, will determine how successfully the algorithm will run. Unfortunately there are no standard rules for determining these values, although a number of methods for estimating suitable values have been proposed (Reeves 1995).

Genetic Algorithms

A Genetic Algorithm (GA) is a form of evolutionary algorithm, based on Darwin's theory of natural selection (Darwin 1859). The algorithm works on the premise that small variations in organisms can accumulate if they induce an increase in the overall fitness and this improves the individuals' ability to reproduce (Reeves and Rowe 2003). When describing GAs, the following concepts are important:

- A **gene** is a single model parameter.
- A **chromosome** is a combination of genes, i.e., a unique model configuration.
- The **population** is the current set of chromosomes that the algorithm is using to explore the parameter space.



Calibration of Simulation Models, Fig. 4 The process of running a genetic algorithm

- **Selection** is the process of choosing the fittest chromosomes.
- **Recombination** is the process of combining the best chromosomes to initialize the new population.
- **Mutation** is a means of exploring a wider area in parameter space by randomly varying some genes in a new population.

The GA is an iterative process, as illustrated by Fig. 4. At each iteration, the algorithm finds the fittest chromosomes (the model configurations with lowest error) and uses these to create the population in the subsequent iteration. The algorithm runs for a set number of iterations or until an acceptable goodness-of-fit value has been found. Although the process of running a GA does not change substantially across implementations, the means of performing selection, recombination, and mutation can be adapted depending on the characteristics of the parameter space. See Reeves and Rowe (2003) for more information.

The main advantage of a GA over the other procedures is that it is able to "home in" on the space with the fittest parameter configurations without devoting effort to exploring suboptimal areas. This makes it much more efficient. Also, mutation allows the algorithm to explore a wider search space which can prevent it from becoming trapped in local maxima.

Calibration of Criminology Models

On the whole, very few simulation models in the field of criminology apply the techniques outlined here. This is partly because simulation modeling has been, until recently, relatively underused in criminology so there is only a limited literature base to begin with. Also, and perhaps more importantly, a large number of studies are *conceptual* rather than *predictive*. Conceptual models do not attempt to replicate real-world crime patterns and instead explore the dynamics of criminology theory in an abstract, artificial environment. Hence it is normally neither necessary nor possible to configure these models to simulate data from the real world. Predictive models, on the other hand, do attempt to replicate real-world patterns and therefore calibration should be an important part of the modeling process.

However, of the limited relevant published studies, very few have applied the formal methods outlined here. For example, Malleson et al. (2013) implement a simulation model of residential burglary and, although automatic calibration is noted as advantageous, the authors result to manual parameter configuration because the extensive run time of the model leads to in insurmountable computational requirements for automated calibration algorithms (Malleson et al. 2012). Similarly, Groff (2007) discusses the

advantages of a "common sense" approach to model evaluation by examining factors such as the degree of clustering and the spatial dynamics of hotspots. Again, formal calibration methods such as those outlined here are not applied, although the authors stress that the field is "wide open" (Groff 2007, p. 99). There are numerous other predictive simulation models that would benefit from a rigorous approach to calibration but, for a number of reasons, calibration is not applied with the same rigor as with simulation models in other fields. Partly, this will be because the methods discussed here are not simple and require a degree of computer literacy. But, also, it is due to the novelty of the models to the field and, with time, it is extremely likely that calibration, as well as the other elements to model evaluation, will form a more substantial part of the modeling process.

Controversies and Open Questions

This entry has made reference to a number of calibration methods that can be used to calibrate simulation models. Other methods are available, including simply adjusting parameters manually. But those discussed here are the most suitable for models of complex systems which often have a large number of parameters, behave nonlinearly, and consume/produce large amounts of data. However, unless reliable observed data can be gathered and methods developed to compare the data to simulation results, the process of calibration will not be able to improve model performance (at least in the sense that the model is a good representation of the real world). The first problem is how to obtain reliable, real-world data on which to build a picture of the underlying system. Police recorded crime data is a common source but has a number of drawbacks:

- Discrepancies in police recording practices mean that some crimes, which are heavily underreported, will be misrepresented in police data – although this is somewhat mediated by the assertion that unreported crime clusters near reported crime (Chainey and Ratcliffe 2005).
- The temporal accuracy is often questionable because the actual time of the event is not always known.
- Spatial accuracy can be variable if a location is hard to code accurately (e.g., somewhere in a park) or if human error corrupts the recorded location.
- Offender data, which can be useful for the calibration of the spatial movements of offenders, by definition only provides information about people who have been in contact with the police and therefore misrepresents the actual population of offenders.

Assuming good-quality crime data are available, it is still not certain that those data are the most suitable for use in calibration. Firstly, many models might be able to recreate the observed crime patterns, but this does not guarantee that any of them correctly represent the internal dynamics of the system. This is known as the identifiability problem and one that is common to all modeling approaches – see, for example, the discussion in Windrum et al. (2007). To determine which, if any, are "correct," it might be necessary to simultaneously calibrate against various different data sources that capture elements of the system other than simply the crimes committed. Examples might include the use of social surveys to represent victim behavior, crowdsourced data to explore "normal" day-to-day behavior patterns or transport data to estimate the routes that people use to navigate cities. The "correct" model will fit the patterns illustrated by these data *and* closely approximate the observed crime data.

Although there is huge scope for improving the calibration of criminology simulation models, relatively little has been done in practice. This is not, however, unexpected. The methods employed are still in their infancy, relative to their traditional mathematical counterparts at least, so it takes some considerable effort to develop a model in the first place. However, as the methods become more widely used and the tools to develop them become easier to manage, there is no reason that standard, widely adopted approaches to calibration cannot emerge.

Related Entries

- ▶ Agent-Based Models to Predict Crime at Places
- ▶ Spatial Models and Network Analysis

Recommended Reading and References

Andresen M, Malleson N (2011) Testing the stability of crime patterns: implications for theory and policy. J Res Crime Delinq 48(1):58–82

Bailey T, Gatrell T (1995) Interactive spatial data analysis. Prentice Hall. http://www.amazon.com/Interactive-Spatial-Analysis-Trevor-Bailey/dp/0582244935

Birks D, Townsley M, Stewart A (2012) Generative explanations of crime: using simulation to test criminological theory. Criminology 50(1):221–254

Chainey S, Ratcliffe J (2005) GIS and crime mapping, 1st edn. Wiley, Chichester

Clark PJ, Evans FC (1954) Distance to nearest neighbour as a measure of spatial relationships in populations. Ecology 35:445–453

Costanza R (1989) Model goodness of fit: a multiple resolution procedure. Ecol Model 47:199–215

Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London

Dray A, Mazerolle L, Perez P, Ritter A (2008) Policing Australia's 'heroin drought': using an agent-based model to simulate alternative outcomes. J Exp Criminol 4(3):267–287

Groff ER (2007) Simulation for theory testing and experimentation: an example using routine activity theory and street robbery. J Quant Criminol 23(2):75–103

Harland K (2008) Journey to learn: geographical mobility and education provision. PhD thesis, School of Geography, University of Leeds, Leeds

Hayslett-McCall K, Qui F, Curtin KM, Chastain B, Schubert J, Carver V (2008) The simulation of the journey to residential burglary. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographic information systems. Information Science Reference, Hershey

Kirkpatrick SS, Gelatt CDC, Vecchi MPM (1983, May) Optimization by simulated annealing. Science 220(4598):671–680

Knudsen DC, Fotheringham AS (1986) Matrix comparison, goodness-of-fit, and spatial interaction modeling. Int Reg Sci Rev 10:127–147

Kuhnert M, Voinov A, Seppelt R (2005, August) Comparing raster map comparison algorithms for spatial modeling and analysis. Photogramm Eng Remote Sens 71(8):975–984

Levine N (2006) Crime mapping and the crimestat program. Geogr Anal 38:41–56

Liu L, Wang X, Eck J, Liang J (2005) Simulating crime events and crime patterns in RA/CA model. In: Wang F (ed) Geographic information systems and crime analysis. Idea Publishing, Reading, pp 197–213

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

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, Leeds. http://www.geog.leeds.ac.uk/fileadmin/downloads/school/research/wpapers/10_1.pdf

Malleson N, See L, Evans A, Heppenstall A (2012) Implementing comprehensive offender behaviour in a realistic agent-based model of burglary. Simul Trans Soc Model Simul Int 88(1):50–71

Malleson N, Heppenstall A, See L, Evans A (2013) Using an agent-based crime simulation to predict the effects of urban regeneration on individual householdburglary risk. Environ Plan B Plan Des 1–28. http://www.envplan.com/abstract.cgi?id=b38057

Metropolis N, Rosenbluth AW, Teller E, Rosenbluth MN, Teller AH (1953) Equation of state calculations by fast computing machines. J Chem Physica 21(6):1087–1091

O'Sullivan D, Unwin DJ (2003) Geographic information analysis. Wiley, Hoboken, NJ

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

Reeves C (1995) Modern heuristic techniques for combinatorial problems. McGraw-Hill, London

Reeves CR, Rowe JE (2003) Genetic algorithms – principles and perspectives: a guide to ga theory. Kluwer, Dordrecht

Rykiel EJJ (1996) Testing ecological models: the meaning of validation. Ecol Model 90:229–244

Windrum P, Fagiolo G, Moneta A (2007) Empirical validation of agent-based models: alternatives and prospects. J Artif Soc Soc Simul 10(2)

Cambridge-Somerville Youth Experiment

Sonia Jain and Alison K. Cohen
Health and Human Development Program, WestEd, Oakland, CA, USA

Synonyms

Cambridge-Somerville youth study