16 Spatial Modeling with GIS

OVERVIEW

- This chapter begins with the necessary definitions and presents a taxonomy of models
- The alternative software environments for modeling are reviewed, along with capabilities for cataloging and sharing models, which are developing rapidly.

LEARNING OBJECTIVES

- **Know what modeling means in the context of GIS;**
- **Be familiar with the important types of models and their applications;**
- **Be familiar with the software environments in which modeling takes place;**
- **Understand the needs of modeling and how these needs are being addressed by current trends in GIS software.**

KEY WORDS AND CONCEPTS

Computational models, geocomputation, spatial and temporal resolution, static models, indicators, cellular models, cartographic modeling, map algebra, model coupling, multicriteria methods, cross-validation, agent-based models (ABM), scripts, uncertainty propagation, sensitivity analysis

OUTLINE

16.1 Introduction

- A clear distinction needs to be made between *data models* and the *spatial models* that are the subject of this chapter.
    - A data model is a template for data, a framework into which specific details of relevant aspects of the Earth's surface can be fitted. It is a statement about how the world looks
    - Models in this chapter are expressions of how the world is believed to *work*, in other words they are expressions of process
- All of the models discussed in this chapter are *digital* or *computational* models, meaning that the operations occur in a computer
    - The term *geocomputation* is often used to describe the application of computational models to geographic problems
- All of the models discussed in this chapter are also *spatial* models. There are two key requirements of such a model:
    1. There is variation across the space being manipulated by the model
    2. The results of modeling change when the locations of objects change
- The level of detail in computational models is measured as *spatial resolution* which is defined as the shortest distance over which change is recorded.
- *Temporal resolution* is defined as the shortest time over which change is recorded
- Spatial and temporal resolution determine
    - What is left out of the model
    - What is the level of uncertainty between the model and the real processes
    - The cost of acquiring data
    - The cost of running the model

16.1.1 Why model?

- To support a design process
- To experiment on a replica of the world

- To examine dynamic outcomes

16.1.2 To analyze or to model?

- Analyzing results in static outcomes
- Modeling can result in dynamic animations

16.2 Types of models

16.2.1 Static models and indicators

- A static model represents a single point in time and typically combines multiple inputs into a single output. There are no time steps or loops.
- USLE is given as an example. Why use GIS to evaluate the USLE?
  - Some of the inputs require GIS for their calculation
  - The inputs and outputs are best expressed, visualized, and used in map form
  - The inputs and outputs are often integrated with other types of data

16.2.2 Individual and aggregate models

- Models of physical systems are forced to adopt aggregate approaches because of the enormous number of individual objects involved, whereas it is much more feasible to model individuals in human systems, or in studies of animal behavior
- Even when modeling the movement of water as a continuous fluid it is still necessary to break the continuum into discrete pieces
  - Some models adopt a raster approach and are commonly called *cellular* models
  - Other models break the world into irregular pieces or polygons
- Models of individuals are often termed *agent-based* models (ABM) or *autonomous agent* models

16.2.3 Cellular models

- Each cell in the raster has a number of possible states which change through time as a result of the application of transition rules.
- Typically the rules are defined over each cell's neighborhood, and determine the outcome of each stage in the simulation based on the cell's state, the states of its neighbors, and the values of cell attributes.
- One of the most important issues in such modeling is calibration and validation

16.2.4 Cartographic modeling and map algebra

- Classifies all GIS transformations of rasters into four basic classes: local, focal, global, and zonal
- Provides a simple language in which to express a model as a script

16.3 Technology for modeling

16.3.1 Operationalizing models in GIS

- Models can be defined as sequences of operations which can be expressed as either graphic flowcharts or *scripts*
- Scripts express models as sequences of commands
- While scripting used to be specific to individual software products, today it is common for scripts to be written in industry-standard languages

16.3.2 Model coupling

- GIS will often fail to provide adequate performance as a modeling environment for very large datasets and large numbers of iterations
- Therefore, spatial modeling is often done by coupling GIS with other software
    - *Loosely coupled* models are run separately with data exchanged as files
    - *Closely coupled* models use the same files
    - *Embedded* models are implemented as GIS scripts or through Graphic User Interfaces (GUI)

16.3.2 Cataloging and sharing models

- Very little effort to date has gone into making it possible to share *process objects*, digital representations of the process of GIS use
- GIServices describe GIS functions being offered by a server for use by any user connected to the Internet
    - Reasons for offering these are described in Section 7.2 and repeated here
    - Examples of services include gazetteer service and geocoding service

16.4 Multicriteria methods

- Are used when there are a number of factors that influence vulnerability and there are different stakeholders with different views about what is important, how that importance should be measured, and how the various important factors should be combined

- Multicriteria decision making (MCDM) is commonly used whenever decisions are controversial
- A common framework is the Analytical Hierarchy Process (AHP), which focuses on capturing each stakeholder's view of the appropriate weights to give each impact factor
    - There weights are inserted as parameters in the spatial model to produce a final result

16.5 Accuracy and validity: testing the model

- Models can often be tested by comparison with past history, by running the model not into the future, but forwards in time from some previous point.
    - But these are often the data used to *calibrate* the model, to determine its parameters and rules, so the same data are not available for testing.
- Many modelers use *cross-validation*, a process in which a subset of data is used for calibration, and the remainder for validating results.
- Models of real-world processes can be validated by experiment, by proving that each component in the model correctly reflects reality.
- However, the outputs of models must always be taken advisedly, bearing in mind a model:
    - May reflect behavior under ideal circumstances and therefore provide a norm against which to compare reality
    - Should not be measured by how closely its results match reality but by how much it reduces uncertainty about the future
    - Is a mechanism for assembling knowledge from a range of sources and presenting conclusions based on that knowledge in readily used form
- Models are subject to uncertainty
- Present in their inputs
    - *Uncertainty propagation* concerns the impacts of input uncertainty on the uncertainty of the outputs
- Present in their parameters
    - *Sensitivity analysis* examines each parameter in turn to see how much influence it has on the results
- Related to the labels used to express the results

16.6 Conclusion

ESSAY TOPICS

1. What do you understand is the scientific meaning of the word ‘model’, and how does this differ, if at all, from everyday use of the word?
2. Give a reasoned definition of the term ‘geocomputation’, and provide examples of its use in GIS analysis.
3. Why model? Illustrate your answer by reference to examples from GIScience.
4. Differentiate between analog and digital models. What are the main difficulties in validating both types of model?
5. What is a cartographic model, and what distinguishes it from other types of model used in GIScience?
6. Compare and contrast ‘modeling’ with ‘analysis’. Why and in what ways do these activities normally interact?
7. With reference to specific examples, explain how dynamic models might be coupled to a GIS. What are the advantages and disadvantages of each approach?
8. You have just completed a climate forecast using a large-scale model of the atmosphere for a situation in which there is a doubling of atmospheric carbon dioxide. The model predicts a global warming over the next 50 years of around 4C. How might you validate this finding?
9. A possible classification of mathematical models is into ‘deterministic’ and ‘stochastic’. Giving examples of each, illustrate what is meant by these terms.
10. A famous meteorologist, Ed Lorenz, once distinguished two types of atmospheric model. The first were designed to predict the future behavior of the present weather for use in weather and climate forecasting, the second to paint ‘what if’ scenarios for how the climate might change when disturbed in some way such as by increased carbon dioxide, nuclear war, deforestation and so on. Of the two types of modeling activity, he was of the opinion that the second is much easier than the first. Can you state why, and give other examples drawn from both human and physical sciences?

MULTIPLE CHOICE QUESTIONS (MCQ)

1. Classify each of the following model as ‘analog’ or ‘digital’
   a) Printed map
   b) Varignon frame

c) Atmospheric general circulation

d) Universal Soil Loss Equation

2. Attempt a definition of the following model properties:
   a) Spatial resolution is ……………………………………………
   b) Temporal resolution is ……………………………………………

3. For each of the following models, classify them by the number of dimensions used to represent space (Point = L^0, Section/Line = L^1, Area= L^2, Volume = L^3) and whether or not there is a time dimension:

Model	**Spatial Dimension?**	**Time?**
Dam Busters model, Application Box 16.1		
Richard Church's evacuation model, Biographical Box 16.2		
DRASTIC groundwater model, Figure 16.7		
Mammoth Cave groundwater vulnerability, Application Box 16.3		
Batty's crowd movement models, Technical Box 16.4		
Conway's Game of Life, Technical Box 16.5		
Keith Clarke's urban growth model, Figure 16.3		
MCD model, Figure 16.15		

4. Draw diagrams to illustrate each of the four basic types of spatial operation in the Map Algebra. These are 'local', 'focal', 'global', and 'zonal'.

5. What does the term analog mean when applied to a scientific model?

Which of the following words best describes analysis and which modeling? STATIC, MULTIPLE STAGED, IMPLEMENTING HYPOTHESES, EXPERIMENTING, MANIPULATING DATA, and GENERATING HYPOTHESES. Write your answers under the headings provided.

Modeling	**Analyzing**

6. Which of the following are valid media for defining a dynamic model:
   a) Computer programs
   b) Wood and glue
   c) Mathematical equations
   d) Flowcharts
   e) Databases

7. In a sentence, say what is meant by each of the following when applied to coupling between a model and a GIS:
   a) Loose ………………………………………
   b) Close ………………………………………
   c) Embedded ………………………………………

8. Define each of the following acronyms used in the Chapter:
   a) ABM
   b) AHP
   c) MCDM
   d) USLE
   e) VBA

ACTIVITIES

1. John R. Gold *et al*. (1992)*Teaching Geography In Higher Education: A Manual of Good Practice*, Chapter 7 has some useful, if dated, materials on computer simulation modeling, which give useful ideas on teaching with, and about, models. They can be used to:
   - provide in class demonstrations;
   - carry out experiments;
   - replicate a known feature or event;
   - relate model behavior and outcomes to the real world;
   - explore alternative realities;
   - provide an environment for problem solving;
   - act as a focus for evaluating geographic theory.

For many of these activities, the portal developed at www.ncgia.ucsb.edu/projects/metadata/web_models.html will be found to be extremely useful.

2. Building a cartographic model. As suggested in the text, Application Box 16.2, visit www.esri.com/news/arcuser/0704/files/modelbuilder.pdf and follow the sequence of steps taken by Rhonda Pfaff and Alan Glennon in using ESRI's *ModelBuilder* ™ to create their model of groundwater vulnerability in the Mammoth Cave watershed.

3. The anatomy of a model. This exercise can be developed using any suitable model that is within the student's discipline. Simply examine the detail of a model, concentrating on aspects such as its formulation in mathematical terms, how it is operationalized, the necessary inputs, what the outputs are used for, and so on. The SOLIM model of A. Xing Zhu at solim.geography.wisc.edu is an accessible example. Examples of other computational models in a GIS frame will be found at www.ncgia.ucsb.edu/projects/metadata/.

4. Model sensitivity to inputs. An important feature of many complex models is their sensitivity to changes in their inputs. This exercise examines this by way of a simple example:

    a) Visit the Dane County website at www.co.dane.wi.us/landconservation/uslepg.htm and examine the materials presented on the Universal Soil Loss Equation (USLE). This has been used many times in GIS to create cartographic models of soil loss over a region. It has as its inputs 'factors' developed from analysis of rainfall, erodibility, slope, crop type, and management. Download the model spreadsheet and experiment to discover the output sensitivity to changing inputs.
    b) Now examine the Temperature Urban Run-Off (TURM) model at the same website. How sensitive are its three outputs to changes in the inputs?
    c) Can you develop indices that would provide a systematic assessment of model sensitivity?
    d) Can you suggest a useful way of measuring model output sensitivity?

5. Models can be classified in many ways. Climatologists, such as Kendal McGuffie and Ann Henderson-Sellers (2005), often classify models by their spatial dimensionality. In a way that will be familiar to GIS students, this starts at the base of a pyramid by point models (L^0) that also do not have time variation and are thus 'equilibrium models'.

Watson and Lovelock's (1985) famous 'Daisyworld' model is of this type. Keeping the same spatial dimensionality and thus treating the Earth-atmosphere system as a single point, Budyko (1969) and Myrup (1969) developed a model that varied its input over time. Myrup's model, designed to teach about urban heat islands, was developed on an analog computer and in digital code by Outcalt (1972). The next step is to make space 1-dimensional (L^1), usually either by looking at variation with height or by latitude (that is, by averaging conditions over longitude), as in a global climate model by Sellers (1973). A two-dimensional spatial model (L^2) without time variation creates what Section 16.2.4 refers to as a cartographic model. Allow time to vary and we have models like the cellular automata discussed in the text. Add a third spatial dimension, the height (L^3) and allow time to vary and you have, for example, a full atmospheric general circulation model (AGCM) of the sort used routinely to predict tomorrow's weather or to experiment in order to investigate global warming.

Either from the literature, or using your imagination, find or suggest examples of every kind of model that arises from a classification of the sort given in the table below:

Dimensionality of space	**Time invariant**	**Time varying**
L^0	?	?
L^1	?	?
L^2	?	?
L^3	?	?

References for this question

Budyko M I(1969 The effect of solar radiation variations on the climate of the Earth, *Tellus*, 21, 611-619.

Kendal McGuffie and A Henderson-Sellers 2005 *A Climate Modelling Primer* , Wiley: NY, 3rd Edition A very comprehensive text with software.

Myrup LO 1969 A numerical model of the urban heat island, *Journal of Applied Meteorology,* 8, 908-910.

Saaty T L 1980 *The Analytical Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York: McGraw-Hill.

Outcalt S I 1972 A numerical surface climate simulator, *Geographical Analysis*, 3, 379-398.

Sellers W D 1973 A new global climate model, *Journal Applied Meteorology*, 12, 241-254

Watson A J, Lovelock J E 1983 Biological homeostasis of the global environment: the parable of Daisyworld, *Tellus*, 35, 285-288. Proof that to influence decisions and debate,

models do not have to be complicated. In terms of its influence on thought, Daisyworld is possible the most influential model of all time.

6. John Conway's *Game of Life* (Technical Box 16.5 and Figure 16.11) is great fun to play, but it also has a serious objective in demonstrating how simple rules, played out over a 'geography', can generate enormously complex behaviors. First, visit the website at www.math.com/students/wonders/life/life.html and run the game itself. Next, consider how this sort of simple model might be used in geography. Examples include Michael Batty's simulation of crowd behavior in crowded spaces, as detailed in Technical Box 16.4. One fruitful area of work has been to develop cellular automaton models of urban land use. In this, each cell had a classified land use, with various rules that at each time step say whether it will change its use or not. What rules do you think might be appropriate for such a model? Roger White and Guy Engelen 2000 have developed some very realistic urban simulations using this approach.

7. Dana Tomlin's Map algebra (see Section 16.2.4 and, for example www.quantdec.com/SYSEN597/GTKAV/section9/map_algebra.htm) is an attempt to define a language that enables cartographic models to be quantified, at least in a raster GIS environment. In this exercise we list extensions to the basic command set:

To introduce time variation and some general operators of use primarily in environmental science modeling using GIS. The best example of this is PcRaster, developed by Willem van Deursen, Cees Wesseling, Peter Burrough, Derek Karssenberg, Edzer Pebesma and Kor de Jong of the Department of Physical Geography, Utrecht University, The Netherlands. There is a very good website at http://pcraster.geo.uu.nl

8. Model user interfaces. Visit www.spatial.maine.edu/~max/MapAlgebraSurvey.pdf and examine the paper by H Thomas Bruns and Max J Egenhofer (1997) User interfaces for map algebra. *Journal of the Urban and Regional Information Systems Association*, 9(1): 44-54. In this they note the importance of interface design, recognizing four basic designs: (a) command line (b) form-based (c) flowchart and (d) 'stack'. Collect as many examples of models as you can, and in each case classify their user interfaces using this approach.

9. The role play exercise takes you through a hypothetical siting study involving a garbage landfill site close to the City of 'Greensea', in which a cartographic model using overlay of a series of GIS products is the basic approach. Although this is technically very easy in a

GIS, the analysis has considerable potential for error in the final locational decision arising from the original data, generalizations in the inputs and outputs, the data pre-processing, and the procedures used.

We first imagine that the City's own GIS team has performed a sieve analysis developing a cartographic model for the decision from four allegedly 'objective' criteria (see below). As a result of this analysis, they have selected a suitably sized site that meets these criteria. Possibly catastrophic for property values and for the inhabitant's well-being, this site is close to the small community known locally as 'East-Central'. East-Central's inhabitants are understandably upset by this decision and are to appeal the decision. As part of their evidence, they wish to discredit the City GIS unit by revealing as many sources of uncertainty as they can in the study. To do this they have hired a newly graduated GIS analyst (you) in which they ask you to:

Produce a short report on the possible sources of error in the final overlay map to be included as part of the community's submission to the City Council

A brief overview of what the City GIS team did in their study and the data sources they used is as follows. Their objective was to site a garbage landfill near to Greensea according to the following four agreed criteria. The site has to be:

Within 250m of the main urban area of Greensea to provide easy access. This layer was found by buffering using the US Bureau of Census City Block data;

Within 100m of a road, for easy access by garbage trucks. This used a layer locally surveyed in 2000 using GPS to record the road center-lines, again developed by buffering these lines.

On slopes of less than 2^0, in order to avoid water-pollution problems with drainage and allow easy working conditions. This was derived from standard USGS 30m digital elevation matrices using the supplied 'slope' command in the GIS;

On land of low agricultural potential. This layer was developed by classification based on the USGS GIRAS series using aerial photography shot in the 1970s and 1980s. At source, the data were projected onto an Albers Equal Area Projection referenced to the 1927 NAD and, to co-register these with the other layers used, they were re-projected onto UTM using NAD

83. Land-cover categories were amalgamated so as to exclude currently productive agricultural land.

These requirements cannot be challenged, since they were set by the City Council at an earlier meeting, and, as typical in this environment, some of the details of what was done are not available to you. However, it is admitted that the analytical strategy the GIS team used was to create a binary (0/1) map for each of these four criteria. On each map areas were coded ‘1’ if, and only if, they met the stated criterion. These four maps were then overlaid so that only those cells that were coded "1" on every one of these maps could be considered as possible sites. In cartographic modeling terms this is a favorability model in which the output *I* isn't a weighted sum of the input *x*, as in Section 16.4, but is the function:

$$I = \prod_{m=1}^{M} w_m f(x_m)$$

In this all the input map values $x_m$are coded 0/1, all the m = 4 layers have a weight w_m of 1.0, there is n additional function (*f*) to generate the inputs, and the capital $\prod$ symbol means take the product of the m values, not as in the text, their summation.

10. Organize a debate on the motion that ‘This house believes that because of its limited data structures, a GIS is an unsuitable framework for any modeling activity’. Almost all of the previous activities are relevant to developing both a case for and a case against this motion.

11. Bidding for a research contract. In teams, develop a bid for a research project to map the risk from landsliding in southern California using GIS techniques together with a dynamical model. The objective is to produce a map at a scale of 1:50,000 of areas so much at risk that they cannot be zoned for housing. The contract, with the Governor's Office of the State can be for up to $500,000, but the work has to be completed within two years. Each team should produce a fully costed and illustrated proposal, specifying its approach and anticipated ‘deliverables’. Two famous Californian slides are documented at

www.geog.ucsb.edu/~jeff/projects/la_conchita/apcg2001_article/apcg2001_article.html and www.ci.fremont.ca.us/Environment/MissionPeakLandslide/Default.htm

California residents will know that the La Conchita site had a second disastrous slide on January 10th in the very wet winter of 2005.

FURTHER READING

Goodchild M F, Parks B O, Steyaert L T editors 1993 *Environmental Modeling with GIS.* New York: Oxford University Press, p. 488

Heuvelink G B M 1998 *Error Propagation in Environmental Modelling with GIS*. London: Taylor and Francis.

Kirkby M J, Naden P S, Burt T P, Butcher D P 1987 *Computer Simulations in Physical Geography,* Wiley: Chichester

Essentially a teaching oriented text with some valuable materials.

Kirkby M, Naden P 1988 The use of simulation models in teaching geomorphology and hydrology, *Journal of Geography in Higher Education,* 12 (1): 31-49

See above!

Marble D F, Anderson B M 1972 *LANDUSE: A Computer Program for Laboratory Use in Economic Geography Courses,* Washington, DC, Association of American Geographers, Technical Paper 8, Commission on College Geography.

A very forward-looking teaching simulation program, now of historical interest, that implemented the 'simple' Von Thunen land use model, but then by relaxing the assumptions rapidly showed how even a simple model could generate seemingly complex land use patterns. The NCGIA develop a similar model for use with IDRISI, see Dodson R F 1991 *VT/GIS: The von Thunen GIS Package*, see http://www.ncgia.ucsb.edu/pubs/publist.html#91-27

Tomlin C D 1990 *Geographic Information Systems and Cartographic Modeling*. Englewood Cliffs, NJ: Prentice Hall.

Unwin D J 1981 Teaching a model-based climatology using energy balance simulation, *Journal of Geography in Higher Education*, 5, 133-138

Describes teaching units in a local context, but has ideas about the educational benefits of modeling.

White R, Engelen G 2000. High Resolution Integrated Modelling of the Spatial Dynamics of urban and Regional Systems. *Computers, Environment, and Urban Systems*, 24: 383-400

Making a simple cellular automaton emulate city growth.

RELATED READING

Maguire D J, Goodchild M F, Rhind D W (eds) 1991 *Geographical Information Systems: Principles and applications*. Harlow, UK: Longman (text available online at www.wiley.co.uk/gis/volumes.html).

24. Spatial data integration, R Flowerdew, pp. 375-87
25. Developing appropriate spatial analysis methods for GIS, S Openshaw, pp. 389-402
26. Spatial decision support systems, P J Densham, pp. 403-12
27. Knowledge-based approaches in GIS, T R Smith and Ye Jiang, pp. 413-25

ONLINE RESOURCES

NCGIA Core Curriculum in GIScience, 2000 (www.ncgia.ucsb.edu/giscc)

2.14.1. Spatial Decision Support Systems (127), *Jacek Malczewski* - *GC notes*

2.14.6. Artificial Neural Networks for Spatial Data Analysis (188), *Suchi Gopal*

NCGIA Core Curriculum in GIS, 1990 (www.ncgia.ucsb.edu/pubs/core.html)

57. Multiple criteria methods
58. Location-allocation on networks
59. Spatial decision support systems
74. Knowledge based techniques