[AM-01-020] Geospatial Analysis and Model Building

Abstract

Spatial modeling is an important instrument to conduct geospatial analysis to understand the world and guide decision-making. In GIS, spatial models are formal languages to express mechanisms of geographic processes and design analytical workflows to understand these processes. With the development of GIS and computer science, various types of spatial models and modeling techniques have become available, which endows the term of “spatial model” with different meanings. This entry provides an overview of common types of spatial models, modeling techniques, and related applications.

Keywords: dynamic models, Model Builder, process models, spatial models

Author & citation

Qiang, Y. (2021). Geospatial Analysis and Model Building. The Geographic Information Science & Technology Body of Knowledge (1st Quarter 2021 Edition), John P. Wilson (Ed.). DOI: 10.22224/gistbok/2021.1.12.

Explanation

1. Introduction
2. Taxonomy of Models
3. Common Types of Spatial Models
4. Techniques for Model Building
5. Closing Remarks

1. Introduction

Spatial modeling is an important instrument to conduct geospatial analysis. In GIS, spatial models combine various analytical tools to derive meaningful information. Different from models used in other disciplines, spatial models involve manipulation of geographic information, and the modeling results vary across geographic space. Spatial models are formal languages to design and implement workflows of geospatial analysis. Additionally, spatial models can represent a digital replica of the world, allowing users to understand the mechanisms of natural and human systems, project future trends, and experiment with what-if scenarios of such systems. With the development of GIS software, spatial modeling can be conducted by calling and combining spatial analysis tools, which releases the burden of programming spatial models from scratch. Given the various meanings and broad applications of spatial models, this entry provides a general introduction of the common types of spatial models, modeling techniques, and their applications.

2. Taxonomy of Models

2.1 Data Model vs. Process Model

The term model has acquired broad meanings and become an overloaded term in the English language. To begin with, a clear distinction needs to be made between "spatial model" as it is being discussed here and "spatial data model." A data model is a conceptual representation of entities and their associations. Data models fill the gap between how the user sees things in the real world and how they are represented and processed in a computer. In GIScience, spatial data models are presentations of spatial primitives and relationships. Common spatial data models include vector data models, raster data models, topological models, and network models, which are discussed in details in the GIS&T Body of Knowledge section on Data Management.

Spatial model as it is discussed in this entry belongs to the category of process models, which represents 1) an expression of how the world is believed to work, or 2) a workflow of spatial operations to achieve certain analytical goals. In the former case, a spatial model formalizes mechanisms of geographic phenomena in computer programs. Such spatial models can represent various social and natural processes such as land cover change, spread of invasive species, and population migration. In the latter case, a spatial model expresses a standard workflow of spatial analysis, which combines various spatial operations to compute meaningful indictors or predictions. Figure 1 (a) shows a simple model process in ArcGIS ModelBuilder, which is a graphic programming interface. Such model processes can be combined into a more complex model as that shown in Figure 1 (b).

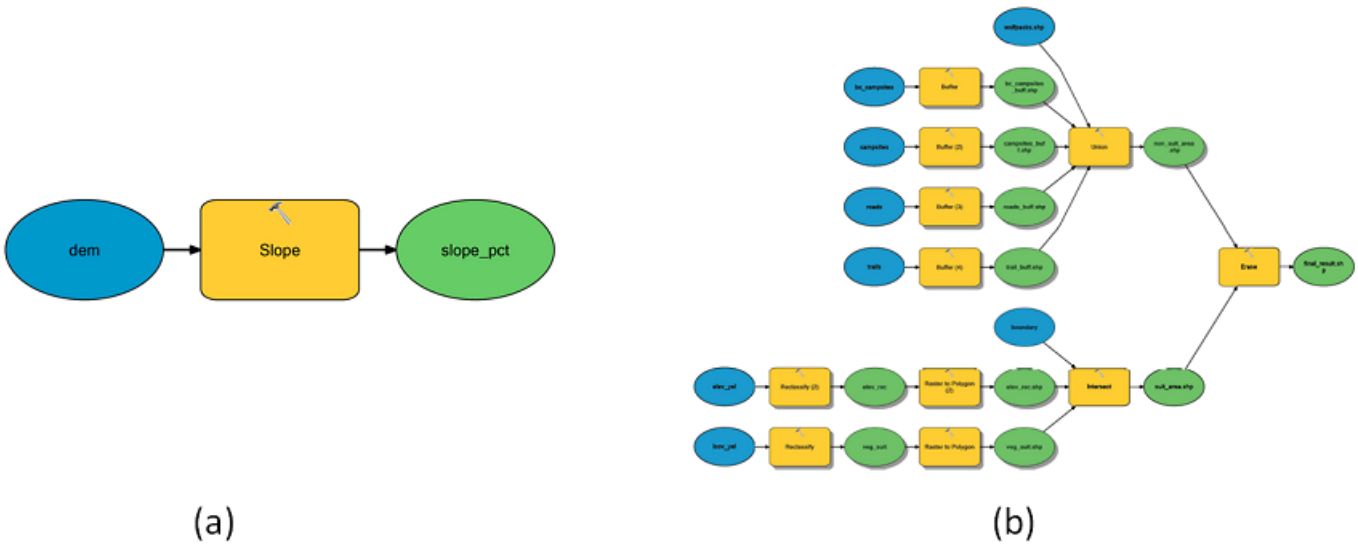


Figure 1. Graphic representation of spatial models in ArcGIS ModelBuilder. (a) a single model process that calculates slope from DEM. (b) a more complex model that consists of multiple analysis processes. Source: author.

2.2 Static Model vs. Dynamic Model

A static model represents processes at a single point of time. Static spatial models are implemented as unidirectional workflows of spatial operations. For example, a landslide risk

model may combine spatial variables of slope, soil type, land cover and precipitation into a risk score expressing the likelihood of landslide over space. The input and output of the model are either temporally implicit or represent a single point of time. On the other hand, dynamic models (also known as spatio-temporal models) represent dynamic processes updating over time. Dynamic models include iterative behaviors and interactions among different elements. **Cellular automata** (CA) and Agent Based Model (ABM) are typical examples of dynamic spatial models.

2.3 Deterministic Model vs. Stochastic Model

A model is deterministic if its output is fully determined by the model without randomness. The same output can be reproduced by repeating the model with the same input and model parameters. Most static models are deterministic and provide a single outcome without consideration of its uncertainty. In contrast, stochastic models include some inherent randomness, which can generate different outputs in multiple runs. In a stochastic model, a distribution of potential outputs can be derived through Monte Carlo simulations. The distribution of outputs indicates the uncertainty of the model prediction. As randomness widely exists in social and natural systems, stochastic spatial models can generate more realistic spatial patterns that include unexpected events. For example, many land use and land cover models include stochastic components to model transition probabilities among different land cover types (Qiang and Lam 2015; Wu 2002).

3. Common Types of Spatial Models

3.1 Spatial Process Model

A spatial process describes the mechanism of how a spatial pattern is formed. The observed patterns of crimes, rainfall and land cover are driven by various spatial processes. A spatial process can be represented in textual description, maps, mathematical equations or computer programs. Static spatial process models quantify spatial patterns (e.g. point pattern or spatial autocorrelation) and make predictions of variables at unsampled locations (e.g. spatial interpolation). Spatial Autoregressive Model and Geographically Weighted Regression (GWR) are of this kind. Dynamic spatial process models describe spatial phenomena evolving in time. Examples of dynamic spatial processes include virus spread, flood formation, and land cover change. At a conceptual level, dynamic spatial processes are described in either the Eulerian view or the Lagrangian view. Eulerian models concern about the change of properties (e.g. temperature, land cover) at fixed locations, while Lagrangian models tracks the movement of objects in space. Among the common spatial models, CA can be considered as a kind of Eulerian model, while ABM falls into the Lagrangian category. Both types of models can be implemented in GIS. The coupling of the two types of models can help to address important problems in coupled natural human systems (Brown et al. 2005).

3.2 Cartographic Model

Cartographic modeling refers to the use of a coordinated set of spatial analysis tools (e.g. buffer, overlay, and reclassification) to solve a spatial problem (Tomlin 2017). Cartographic models are often represented in graphic forms like a flowchart (e.g. Figure 1). Most cartographic models are temporally static as they include spatial variables at a fixed point

in time. Suitability modeling is a typical application of a cartographic model, which assigns ranks or scores to the land to express its utility for a specific use. Figure 2 illustrates a suitability model of finding suitable sites for a new school. The preferred school sites are places where slope is gentle, close to recreational sites, and far from existing schools. The model represents the three criteria in spatial data and combines them into a map expressing the suitability for the school site. The value of a cartographic model is twofold. First, a cartographic model presents a graphic representation of a spatial analysis workflow, which facilitates model comprehension and sharing. Second, cartographic models can be automated in GIS for repetitive modeling work and batch data processing. Cartographic models can be developed using both Python scripting or ArcGIS ModelBuilder.

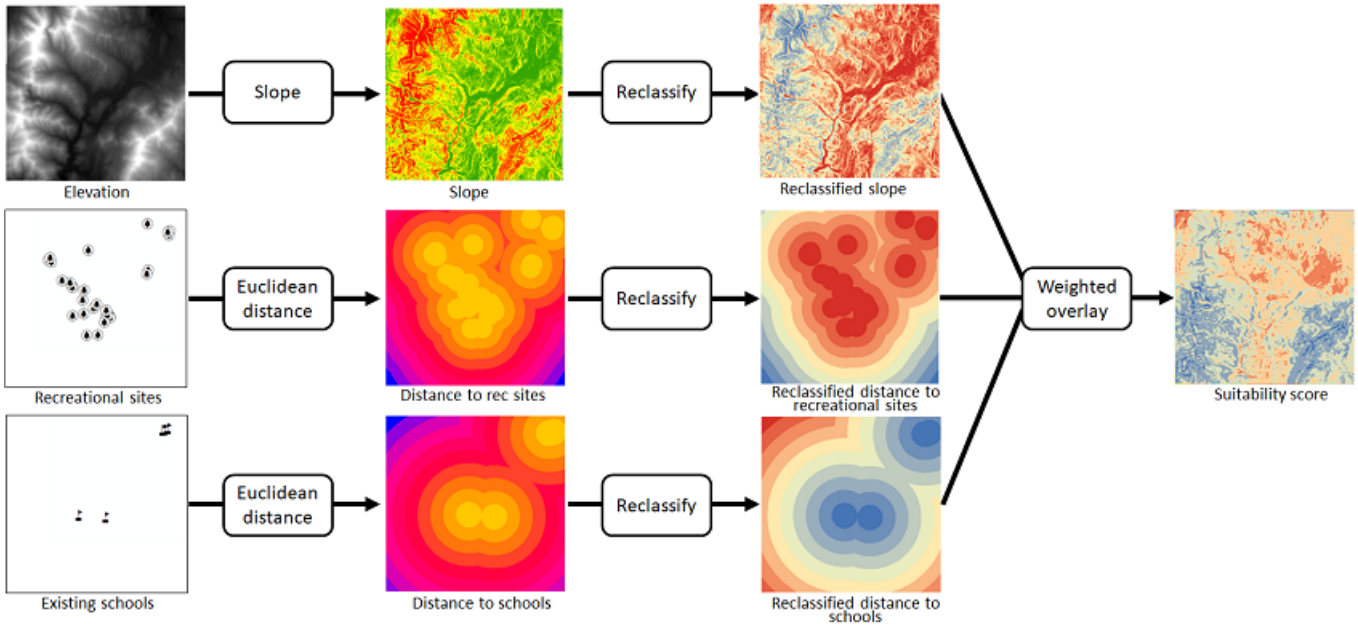


Figure 2. The cartographic model of a suitability analysis. Source: author.

3.3 Cellular Automata

Cellular automata (CA) is a modeling technique widely used to represent dynamic processes. CA represents space as a regular grid of cells, where each cell has a number of pre-defined states. The state of a cell updates in each time step (iteration) according to its neighborhood condition. As a classic example of CA, Conway's Game of Life assumes binary states (i.e. life or dead) of discrete cell in a regular grid, where the state of each cell updates through iterations based on simple rules of its neighbors. More details of CA and Game of Life can be found in the GIS&T BoK entry on Cellular Automata. In a GIS, Conway's Game of Life can be implemented in a spatial model that consists of raster analysis tools. Different from the unidirectional workflow of static models (e.g. the suitability model in Figure 2), CA includes feedback loops where the model output is used as input in the next step (e.g., Figure 3). In a CA model, time is expressed as iterations that defines the relative order of the model states. Due to the spatial nature of CA, GIS provides a powerful platform for building CA models. Currently, CA has been widely applied to model geographical processes such as land cover change, urban growth and invasive species expansion.

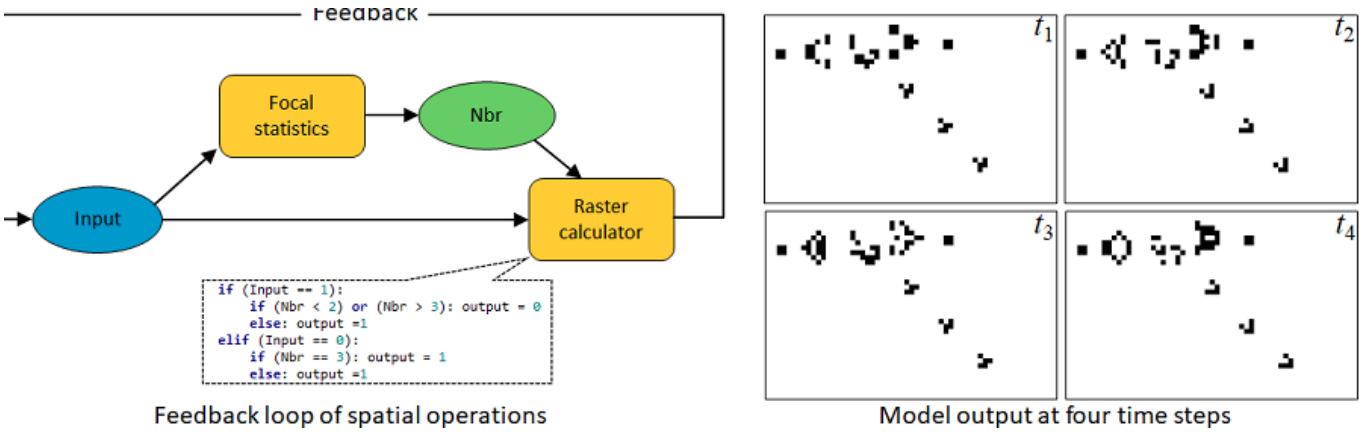


Figure 3. Representing Conway's Game of Life in a spatial model. (a) The model processes in ModelBuilder. (b) Updates of the CA at different time steps. Source: author.

3.4 Agent Based Models

Agent-based models (ABM) represent a dynamic system as a collection of autonomous decision-making agents in a virtual environment. Each agent individually assesses its situation and makes decisions based on a set of behavior rules. Similar to CA, an ABM system updates at discrete time points and forms a repetitive feedback loop. Agent behavior rules are the central to the design of ABM. The common features of the rules include autonomy, heterogeneity, activity, bounded rationality, and mobility. As the agent-level behaviors cannot be modeled in common spatial analysis tools, ABM models are usually developed in non-GIS platforms (e.g. Repast and NetLogo), where GIS data can be plugged in to model geographic components.

4. Techniques for Model Building

4.1 Multicriteria Decision-Making Analysis

The integration of GIS and Multicriteria Decision-Making Analysis (MCDA) can solve decision-making problems involving multiple criteria and constraints in a geographic space. GIS-based MCDA represents evaluation criteria in spatial variables and combines the variables into a final map to facilitate decision making. Most GIS-based MCDA use a linear weighted equation (e.g. Equation 1) to combine the spatial variables. In Equation 1, xi denotes the criteria and wi represents the corresponding weights of the criteria. The selection, scaling and weighting of the criteria are determined by modelers' personal opinions or experts' voting. Then, spatial overlay tools are used to combine the criteria into a spatial indicator y. Suitability analysis (e.g. Figure 2) is a typical example of MCDA. Additionally, MCDA is also used for spatial assessment of social vulnerability, disaster risk and public health. However, MCDA is often criticized for the arbitrary criteria selection and weighting, lack of empirical validation, and its inability of handling complex and non-linear relations.

$$y = \sum_{i=1}^{n} w_i f(x_i)$$

Equation 1.

4.2 Statistical Methods

A statistical model can be used to derive the equation of spatial models if empirical data are available. A statistical model trained from a sample dataset can be extrapolated to predict indicators at unsampled locations. For instance, Glenz et al. (2001) uses multivariate logistic regression to calibrate the coefficients (weights) in a suitability model for wolf habitat. The model is derived from 143 sampling locations where the regular wolf presence or absence was recorded. The model can predict the probability of wolf presence (indicating wolf habitat suitability) in the entire study area using socio-environmental variables extracted from GIS data. As another example, Lam et al. (2016) utilized discriminant analysis to build relationships between community resilience with socio-economic conditions of communities. The model is trained using empirical disaster data, and then is applied to predict community resilience in potential disasters. All these models rely on GIS tools to convert various formats of geospatial data into uniform variables for statistical modeling. The statistical approach overcomes the issues of arbitrary parameters (e.g. weights) in the MDCA models. The derived models are explanatory (white-box), as model coefficients indicate the effects of input variables to the output variable. However, the statistical models are still based on linear equations, which are unable to handle non-linear relations and feedback loops in complex systems.

4.3 Artificial Intelligence

In recent years, artificial intelligence (AI) techniques have been widely applied in spatial modeling. Unlike the statistical models, AI does not assume linear relations in the data and thus can model complex processes in natural and social systems. Commonly used AI algorithms include support vector machine (SVM), decision-trees, genetic algorithm (GA) and artificial neural network (ANN), which can address classification or regression problems. Particularly, deep neural network (DNN), which is a special type of ANN, has gained great success in spatial modeling due to its outstanding performance in handling high-dimensional data (e.g. images and 3D datasets). With the increasing applications of AI in geography, GeoAI has emerged as an interdisciplinary area that receives much of attention from both the geography and computer science communities. These methods are discussed with details in the GIS&T BoK entry on Artificial Intelligence Approaches. Compared with the statistical methods, AI is advantageous in prediction accuracy, automated training process, and the ability of handling non-linear and complex relations.

Figure 4 illustrates the procedure of applying AI in land cover change modeling (Qiang and Lam 2015). In this study, an ANN was trained to associate socio-economic variables (input variables) with empirical land cover changes (target variable). The derived ANN was then implemented in the transition function in a CA to simulate land cover changes. The ANN predicts the probabilities of a land cell staying unchanged or changing to other land cover types. Based on the probabilities, a stochastic component determines the land cover type at the next time step. This process repeats through multiple iterations until a threshold of change quantity is reached. This model was used to project future trends of urban growth,

land loss, and ecological changes in a vulnerable coastal area in the Mississippi Delta.

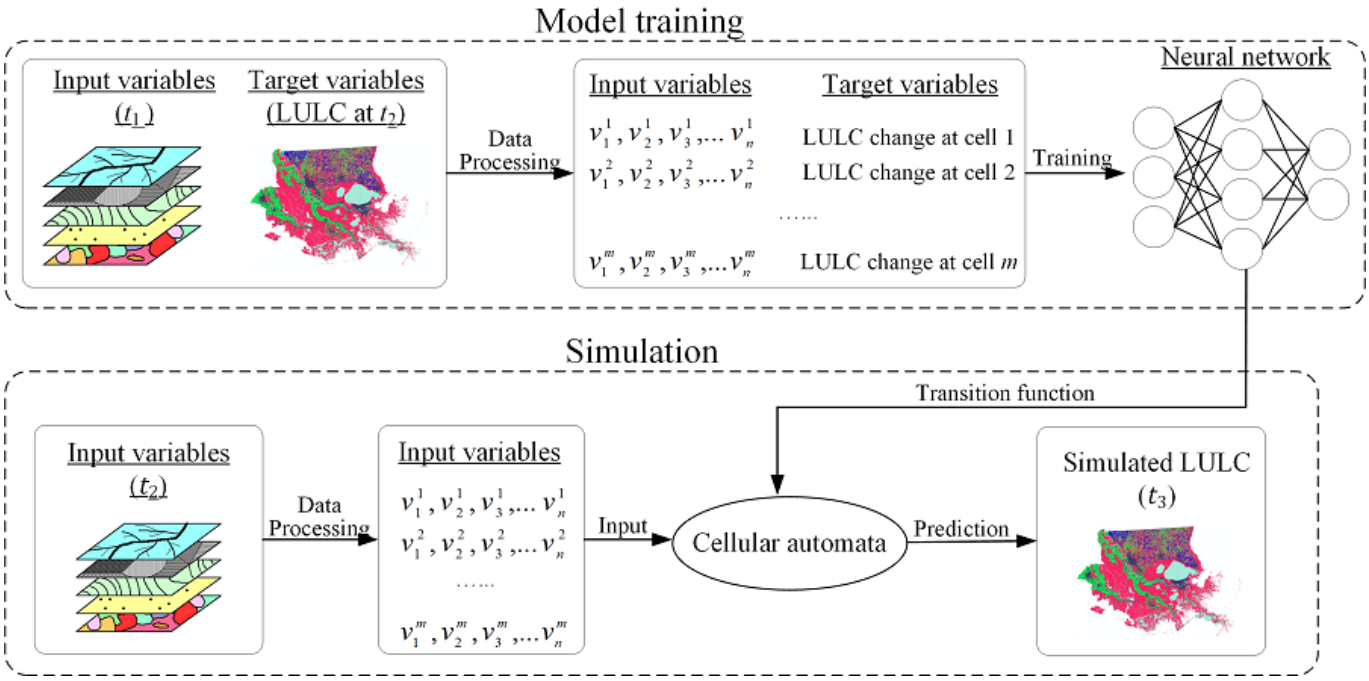


Figure 4: the process of modeling land cover changes using artificial neural network (ANN) and Cellular Automata (CA). Source: author.

4.4 Model Validation

Validation is an important step in spatial modeling work. Unfortunately, model validation is often neglected in scientific research because 1) the predicted reality is not yet available, and/or 2) people tend to trust the results of computer models built by experts in the field. Ideally, model predictions need to be compared with real conditions to assess the model validity and accuracy, regardless of the modeling techniques and modelers' expertise. A common validation method is testing whether a model can accurately predict past scenarios before applying it to project the future. Alternatively, a model can be validated by testing its prediction in another dataset or study area that is different from the dataset or study area where the model was trained. Even within the same dataset, cross-validation can be conducted by separating the data into two sets: one set is for model training and the other is for data validation. The accuracy of the model prediction can be expressed in various statistics such as Root Mean Square Error (RMSE), recall and precision, Kappa coefficients and confusion matrix. In addition to the validation of prediction accuracy, analyses of uncertainty and sensitivity are also important steps to prove a model's validity and usefulness.

5. Closing Remarks

A spatial model is an important instrument to convert geospatial data into meaningful and useful information for understanding the world and making decisions. The advances of GIS and standardized geospatial data have greatly eased the process of spatial modeling. This

entry provided a general introduction to terminology, common techniques and applications of spatial models.

References

Brown, D. G., Riolo, R., Robinson, D. T., North, M., and Rand, W. (2005). Spatial Process and Data Models: Toward Integration of Agent-Based Models and GIS. Journal of Geographical Systems 7(1):25–47.

Glenz, C., Massolo, A., Kuonen, D., and Schlaepfer, R. (2001). A Wolf Habitat Suitability Prediction Study in Valais (Switzerland). Landscape and Urban Planning 55(1):55–65.

Lam, N. S. N., Reams, M., Li, K., Li, C., and Mata, L. P. (2016). Measuring Community Resilience to Coastal Hazards along the Northern Gulf of Mexico. Natural Hazards Review 17(1):e193–e193.

Qiang, Y., and Lam, N. S. N. (2015). Modeling Land Use and Land Cover Changes in a Vulnerable Coastal Region Using Artificial Neural Networks and Cellular Automata. Environmental Monitoring and Assessment 187(3):57.

Tomlin, C.D. (2017). Cartographic Modeling. In International Encyclopedia of Geography: People, the Earth, Environment and Technology (eds D. Richardson, N. Castree, M.F. Goodchild, A. Kobayashi, W. Liu and R.A. Marston).

Wu, F. (2002). Calibration of Stochastic Cellular Automata: The Application to Rural-Urban Land Conversions. International Journal of Geographical Information Science 16(8):795–818.