[AM-06-081] GIS-Based Computational Modeling

Abstract

GIS-based computational models are explored. While models vary immensely across disciplines and specialties, the focus is on models that simulate and forecast geographical systems and processes in time and space. The degree and means of integration of the many different models with GIS are covered, and the critical phases of modeling: design, implementation, calibration, sensitivity analysis, validation and error analysis are introduced. The use of models in simulations, an important purpose for implementing models within or outside of GIS, is discussed and the context of scenario-based planning explained. To conclude, a survey of model types is presented, with their application methods and some examples, and the goals of modeling are discussed.

Keywords: calibration, forecasting, modeling, scenario-based, sensitivity analysis, simulation

Author & citation

Clarke, K. (2019). GIS-Based Computational Modeling. The Geographic Information Science & Technology Body of Knowledge (3rd Quarter 2019 Edition), John P. Wilson (ed.).
DOI: 10.22224/gistbok/2019.3.6

Explanation

1. Definitions
2. Context
3. Model Coupling and GIS
4. Model calibration
5. Simulation
6. Types of Models
7. Conclusion

1. Definitions

computational model: a numerical simulation based on a model

model: a system of postulates, data, and inferences presented as a mathematical description of an entity or state of affairs

scenario: an account or synopsis of a possible course of action or events

sensitivity analysis: the study of how the uncertainty in the output of a mathematical model or system can be divided and allocated to different sources of uncertainty in its inputs.

simulation: the imitative representation of the functioning of one system or process by means of the functioning of another for examination of a problem often not subject to direct

experimentation

uncertainty: lack of sureness about something

2. Context

A **model** is a simplified abstraction of a system whose purpose is to provide information about the functioning of the system, including future states of the system. The term model is applied in GIScience to data, in that geographical data are assumed to represent measurements of objects or features that can be abstracted as points, lines, areas, fields, etc. How accurately, both geometrically and ontologically, the features are captured by the measurements that become the object, very much defines the accuracy, precision, **uncertainty** and level of confidence that we place in the **computational model** and its derivatives. The term model can also apply to GIS tools that manipulate geographical data and create transformations. For example, a dataset can be geometrically manipulated to change the geodetic model on which the data are mapped. Similarly, classes within the data can be aggregated to make a more generalized map model of the landscape, with fewer land cover types for example. When we change data structures, say converting from raster to vector data, such transformations are inevitable because the ontology and assumptions of the data model must be reexamined to make basic decisions about how the data are treated computationally.

A broader use of the term model is that of building a **simulation** of a system that takes data as inputs, modifies the data as part of a process, and then creates possible alternate or future states of the system. For example, within a GIS a typical modeling exercise might involve reading a digital elevation model (DEM), making assumptions about scale, resolution, algorithms and process to create a model of downslope flow accumulation, then thresholding the values to create a connected stream network. Note that this is fundamentally different from mapping streams from imagery or field data, the modeled stream network is a product of the model itself, and would exist even if the mapped area had no surface water flow at all. A future state could then be simulated by making the water flow proportional to rainfall, then using a rainfall quantity above the normal range. In this case we simulate an unknown or even highly unlikely future state. Nevertheless, the model still has purpose to explore, inform, educate or estimate.

3. Model Coupling and GIS

There is a broad range in the linkages between GIS and computational models. At one extreme, the model can function completely independently of the GIS, and the role of GIS is simply to manage or preprocess spatial data for the input to the model, or to display or further analyze output. This is called loose coupling, and is a common strategy in computational models where performance is an issue, or where the GIS tools are inadequate for the modeling task. At the opposite extreme lie fully coupled models in which the entire model and its functions are built using the GIS tools or scripting language available within the GIS software. For example, the IDRISI TerrSet GIS contains modules for surface analysis, change and time series analysis, and programming tools for model development and deployment. An additional trend is for GIS models to be buildable in common open-source environments that also support statistics, visual analytics, web deployment and methodologies, for example the combination of R, R-studio, R-Shiny and

other R analytical libraries.

Many GIS computational models share tools and techniques for conducting the sequence of operations that constitute modeling. A model must initially be designed, and several environments exist primarily for the purpose of building, modifying and testing new models. Examples are NetLogo, ArcGIS Model Builder, and Anylogic. Models in these systems are often shared through common libraries or demonstration use cases, and can be used in instruction and education, both in model construction and in use and testing. After design, a model is implemented. Implementation involves choosing a computational environment, operating system, set of libraries or programming language. The increasing availability of modules for common methods or functions—such as classification, machine learning, visual display and interaction—give open source contributory modeling tools such as R a distinct advantage, especially when they can interact directly with open source GIS tools. Many models require considerable CPU time, and so require the ability to use Graphical Processing Units, parallel computers, clusters and virtual machine methods. Models that use fine spatial or temporal resolution, create large numbers of outputs such as Monte Carlo simulations, or that use computational solutions such as genetic algorithms, machine learning, or brute force processing are candidates for these computational acceleration methods.

4. Model calibration

Once operational, a necessary initial step is calibration. Many factors are inherent to a model, such as the values of constants and multipliers, and are assigned using assumptions, published values, modeling standards, or common sense. Some of these constants have direct implications for computation that are amplified by numerical and computational complexity. For example, pseudo-random numbers have repeat cycles that are commonly ignored, and conversion factors that work well with small numbers are assumed to work for large also (the feet to meter conversion factor, for example, is often assumed to be 0.3048, but is in fact 0.304 800 609 601 2). Some key model constants are set by trial and error, or by empirical experimentation known as mono-looping, in which the model is repeated for increments of a single value, while the remaining values are held constant. A second level of calibration involves executing a model while repeatedly comparing its output for a known state or time with data attained independently. For example, hind-casting involves calibrating a model by matching its performance against a known set of control data, to select those parameters or coefficients that have the highest degree of explanation, matching or fitness. This can be facilitated by methods such as genetic algorithms, landscape metrics matching, or simulated annealing. The SLEUTH land use change model, for example, has used both brute force (trying all possible values) and a genetic algorithm trained using a compound fitness metric. The set of methods for measuring model fit to reality is rich, and includes regression, shape matching, a set of Kappa and other standard error statistics, information metrics, receiver operator characteristic and area under the curve and the likelihood or odds ratio. Many measures make use of the confusion or contingency matrix, which tracks errors of omission and commission aspatially by class.

A penultimate level of modeling is to conduct a **sensitivity analysis**. Sensitivity analysis explores the interconnectedness of the inputs, parameters and model processes with the goal of seeing how important each is in determining the robustness of the model to random or systematic error. For example, the seed value for a pseudo-random number generator

could be set across a large range of settings, and the model outputs compared, using the same set of methods available for calibration; or a DEM-derived stream network created at multiple spatial resolutions compared to determine which gives the best results. Each of the many input and model constant factors could make a model's output unstable, or factors that are insignificant may add no explanatory power compared to others, and so can be eliminated from the model.

5. Simulation

Once fully calibrated and tested, a model can be used for simulation. Simulations are often about the nature of an unknown future, but simulations can also consist of interpolation at unknown locations or points in time, across different spatial scales, resolutions or extents, or at limits untestable in the real systems. Examples are evacuation models, traffic flow models and cascading failure models. When simulations do involve the future, scenario-based modeling is usually used when models are involved in planning (Xiang & Clarke, 2003). In scenario-based modeling, parameters and data are targeted toward particular outcomes, sometimes to show best and worst cases, but also to explore options that can be related to policy or outcomes. Examples are the International Geosphere-Biosphere Programme **scenarios** for global climate change (Carter et al., 2001). Many such models are explanatory in nature, that is, the model's forecasts allow experiments about the system the model represents that increase our scientific understanding sufficient to allow intervention or improved design. Such experimentation may be impossible or unethical without computational modeling. For example, earthquake models can be used to explore building failures and to hopefully reduce the risk of damage or collapse. Modeling landslides can inform land use planning and zoning, and measure the risk of landslides at a particular location.

Ultimately a model can be validated. In validation, a forecast or simulation is compared to an actual outcome of the system, so that the model's value and effectiveness can be measured. For example, a voting model can be tested against actual election results. This allows selection of one model versus another, a meta-assessment of the model's effectiveness, or participation in multi or linked modeling. Linked models allow one model's outputs to become another's inputs. Multi-modeling compares model outcomes to see where there is an agreement in behavior or forecasts that is consistent across several models. Climate change modeling, for example, often conflates the results of ten or more models derived and run independently of each other (Flato et al., 2013).

6. Types of Models

The range of computational models enabled by GIS is extreme, and covers all aspects of geographical science from physical to human. Here we except those models that are abstract and conceptual, often termed a "framework," and intended to guide thought and problem formation rather than to perform computations and simulations. Simple models are often formulas that relate one variable to another, or to geographical coordinates, such as the gravity model or basic models of spatial interaction or potential. Curve fitting and regression methods are often used to reveal the structure of the formula. Goodchild (2005) noted that models make a dichotomy of approaches to space, either assuming it to be a continuous abstract phenomenon like a Euclidean plane, or a set of discrete point, lines, areas or objects about which measurements are made. In either case, GIS-based computational models can model either static or dynamic phenomena. The traditional GIS

model of a discrete geometry determining a two-dimensional bounded space, such as a nation or US state, essentially holds the space constant, and conducts analysis of the properties of the space such as distance decay, spatial autocorrelation, correlation among attributes or degree of randomness or clustering. Many such models use multivariate and logistic regression, network analysis, optimization methods, linear programming, correlation analysis, or multi-criterion analysis. Models are used to link variables, assign locations and facilities or services, weight criteria across variables and to interpolate values where data are unavailable.

On the other hand, an increasing number of GIS-based computational models are dynamic, that is they apply spatial principles across both space and time. At the simplest, models can build equations that determine change, and apply them to modify spatial distributions into the future (or past), a type of model known as systems dynamics (Richardson, 2011). For example, least energy travel paths have been applied to foot travel among the Wari civilization in Peru (Covey et al., 2013). Other model types build change into the model structure, and use observations about a distribution at different points in time to derive rules, behaviors or trajectories that can be used in simulation. These models use agent based modeling (ABM), cellular automata (CA) (Clarke, 2013), Monte Carlo simulation, stochastic simulation, Markov models, and numerical differential equations. An example of a cellular automaton model that combines Monte Carlo and Markov methods with computational calibration methods is SLEUTH, while Dinamica EGO is both a land use change model and an environmental modeling platform created to explore cellular automata and systems dynamics (Pérez-Vega et al., 2012).

Dynamic models like cellular automata and ABM are examples of complex systems models (Holland, 1998). Complex systems models are known for their ability to simulate non-linear behavior, where small changes can initiate subtle and transformative state changes that can lead to stability, randomness, or complexity in an overall system. These models use basic elements (cells in CA, agents in ABM) who follow simple local rules that when aggregated can create emergent unpredictable behavior. Such models use ABM, CA, difference equations and non-static Markov chains to simulate future states and distributions.

Another way to classify models is as stochastic versus deterministic. Deterministic models produce a single result or future state set. Models can also use repeat execution, random perturbation and the addition of noise not just as an aid in sensitivity analysis, but as a way to compute and communicate uncertainties and the robustness of models. Stochastic models can offer not just a forecast state, but also the probability of that state being the outcome. Techniques employed in these models include stochastic and conditional simulation, logistic and multiple regression, fuzzy set theory, Weights of Evidence analysis, classification methods, Monte Carlo simulation and more routine probability and statistics. As an example Yilmaz (2010) compared different modeling methods for landslide susceptibility, while Zhu et al. (2001) examined models for soil mapping.

Social science models often involve equations, constants or behaviors learned or extracted from observed human behavior. Examples are travel behavior learned from travel diaries, vehicle tracks or surveys (Wilkie at al., 2012) , or farmer preferences for decision-making, learned by mixed methods, field surveys and interviews, and then built into systems dynamics or agent based models. The now ubiquitous availability of social media and point-of-sale business transaction data have led to new sets of methods being integrated into the

GIS-computational modeling toolkit (Tsou, 2015).

A final division of models is by their purpose. Obviously different models suit different geographical domains, phenomena and processes, however many GIS-based computational methods and approaches share components in common, such as techniques, user-interface tools and geocomputational methods. Models are used to help understand a system by its simplification; to educate students or professionals about how a process operates; to produce realistic or extreme future states based on scenarios and assumptions; and to anticipate the outcome of either seeing a process continue, or intervening in the process by experiment, planning or policy. Models used in planning or important public decision-making need to be easy to understand (at least in principle) and transparent. They should depend on reliable, publicly available data that is available for testing and experimentation. However, they should produce accurate and meaningful forecasts and interpolations that inform their users and those impacted by the systems in question (Clarke, 2013; Batty & Torrens, 2001).

7. Conclusion

In this essay, GIS-based computational models have been examined. Modeling is a complex field, and here the focus was on models that simulate and forecast geographical systems in time and space. The discussion covered the integration of the many different models with GIS. It also discussed the critical phases of modeling: design, implementation, calibration, sensitivity analysis, validation and error analysis. The use of models in simulations, an important purpose for implementing models within or outside of GIS, was discussed and scenario-based planning introduced. Lastly, a survey of model types, with their application methods and some examples was presented.

Models ultimately have value in their ability to explain past observations and processes, and to predict future observations and conditions with accuracy and repeatability. Model complexity and tool use, especially within a GIS, place constraints on modeling in terms of the costs and time spent building and using models. Simplicity is often a worthy goal, and facilitates model acceptance and ease of use, but at the cost of explanatory power. The ability of a model to integrate, either tightly or loosely, with other models and analyses is often important in model choice. Lastly, to be effective, a model should be well calibrated so as to best resemble the system it simulates, and tested for its sensitivity to noise, error, non-linearities and unrelated variables. Lastly, model results should be replicable by others. Often models are applied by specialists or consultants, and merely the results shared. Open source models and shared code can go a long way toward assuring this last factor.

References

Batty, M., & Torrens, P. (2001). Modeling complexity: the limits to prediction. (CASA Working Papers 36). Centre for Advanced Spatial Analysis: London, UK.

Carter, T. R., & La Rovere, E. L. (2001). Developing and Applying Scenarios. In J. J. McCarthy, O. F. Canziani, N. A. Leary, D. J. Dokken, K. S. White (Eds.), Climate Change 2001: Impacts, Adaptation and Vulnerability, Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge and New York: Cambridge University Press.

Clarke, K. C. (2014). Why simulate cities? GeoJournal, 79(2), 129-132.

Clarke, K. C. (2018). Cellular Automata and Agent Based Models. In M. M. Fischer and P. Nijkamp (Eds.), Handbook of Regional Science. 2ed. Springer-Verlag GmbH Germany, part of Springer Nature.

Covey, R. A., Bauer, B. S., Bélisle, V., & Tsesmeli, L. (2013). Regional perspectives on Wari state influence in Cusco, Peru (c. AD 600–1000). Journal of Anthropological Archaeology, 32(4), 538-552.

Crosetto, M., & Tarantola, S. (2001). Uncertainty and sensitivity analysis: tools for GIS-based model implementation. International Journal of Geographical Information Science, 15(5), 415-437.

Flato, G., Marotzke, J., Abiodun, B., et al. (2013). Evaluation of Climate Models. In Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Doschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley, Eds., Cambridge University Press, pp. 741-882.

Goodchild, M. F. (2005). GIS and modeling overview. In D. J. Maguire, M. Batty, and M. F. Goodchild (Eds.), GIS, Spatial Analysis, and Modeling. Redlands, CA: ESRI Press. pp 1-18.

Holland, J. H. (1998). Emergence: From Chaos to Order. Redwood City, California: Addison-Wesley.

Pérez-Vega, A., Mas, J.-F., & Ligmann-Zielinska, A. (2012). Comparing two approaches to land use/cover change modeling and their implications for the assessment of biodiversity loss in a deciduous tropical forest. Environmental Modelling & Software, 29(1), 11-23.

Richardson, G. P. (2011). System Dynamics. In S. Gass and C. Harris (Eds.) Encyclopedia of Operations Research and Management Science. Kluwer Academic Publishers.

Sui, D. Z. (1998). GIS-based urban modelling: practices, problems, and prospects. International Journal of Geographical Information Science, 12(7), 651-671.

Tsou, M.-H. (2015). Research challenges and opportunities in mapping social media and Big Data. Cartography and Geographic Information Science, 42(1), 70-74.

Wilkie, D., Sewall, J., & Lin, M.-C. (2012). Transforming GIS Data into Functional Road Models for Large-Scale Traffic Simulation. IEEE Transactions On Visualization and Computer Graphics, 18(6), 890-901.

Xiang, W-N, & Clarke, K. C. (2003). The Use of Scenarios in Land-Use Planning. Environment and Planning B: Planning and Design, 30, 885-909.

Yilmaz, I. (2010). Comparison of landslide susceptibility mapping methodologies for

Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environmental Earth Sciences, 61(4), 821-836.

Zhu, A. X., Hudson, B., Burt, J., Lubich, K., & Simonson, D. (2001). Soil Mapping Using GIS, Expert Knowledge, and Fuzzy Logic. Soil Science Society of America Journal, 65(5), 1463-1472.