Paper presented at the 3rd AGILE Conference on Geographic Information Science in Helsinki, May 25-27, 2000.

Modeling with GIS: A game of chess

Thomas Gumbricht

Birger Jarlsgatan 102 b, 114 20 Stockholm, tel. +46 (0)8 16 60 02,
Post doc at University of Witwatersrand, Dept of Geology,
Private bag 3, 2050 Wits, South Africa
thomas_gumbricht@hotmail.com

Abstract. Modeling processes in GIS is both a step forwards and a step backwards. The article analyses some of the shortcomings in GIS modeling using the game of chess as an analogue for spatial processes. A research agenda for improved dynamic modeling in GIS is concluded.

1. Introduction

The dichotomy between data focusing, static and two dimensional (2D) geographical information systems (GIS) on one side and dynamic, mathematical environmental models on the other have been discussed by several authors (Fedra, 1993; Burrough et al., 1996, Maidment, 1996, Fotheringham, 2000). Since Newton published his Principia in the 17th century mathematical modeling of physical phenomena has been very successful. That models based on Newtonian cause-effect relations are at best approximations has been revealed during the last century. Elaborating the models of physical (non living, reversible phenomena) to account for biological (living, non-reversible) phenomena poses great challenges. It is thus not per se evident that the most elaborate physical model will be the most tractable for environmental applications.

A model can have one of two purposes, to improve understanding or to reproduce and forecast particular phenomena. The quality of a model can hence be defined as its ability to represent the real world phenomena studied, as well as how it fulfils its intended purpose, including the users ability to comprehend the model and its output.

This article aims at highlighting some of the weaknesses with environmental modeling in a GIS environment; a research agenda for overcoming these weaknesses is concluded. The game of chess is used as an analogy – the reader should bear in mind that all analogies have weaknesses, but so have models.

2. A simple playing arena – rectangular grids of square

It is seldom convenient to have a sphere for portraying the earth's surface; hence humans have used 2 dimensional (2D) maps for at least 5000 years. As most GIS are 2D, spatial entities in a GIS are represented in what basically comes down to either raster (continuous fields, natural phenomena) or vector (discrete objects, anthropocentric views and phenomena) data models (Goodchild, 1993). For simplistic reasons, the raster model (fig. 1) is often preferred in modeling.

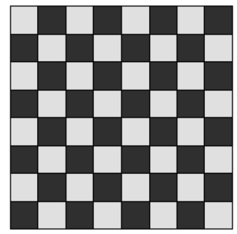
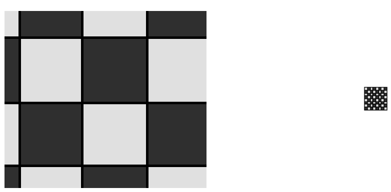


Fig 1 A raster

The raster landscape in figure 1 is also the landscape where the game of chess takes part. Let us assume that we are unaware of the game of chess. Once we have identified the problem from a GIS perspective we must decide on a data model to use and how to capture the data. The chessboard can be captured as primary data or from an existing analogue source. Spatial data is sampled at a certain scale (grain size) and time (Fig. 2). This information is frequently lost on the way to the end user



F*ig. 2 Two raster data sets with different grain size.*

For the sake of argument let us assume that we have some *a priori* knowledge of the arena we are observing. We thus decide on a stratified sampling of data. For each square of unit distance we take one sample at a randomized point. Let us further assume that there is neither an error in position, nor in the obtained value (*sic!*). Rasterising the landscape by honoring the observed value for each unit square (whether via an interpolation method or direct rasterisation does not matter) gives the correct (as we happen to know in this simple case) landscape (fig. 3).

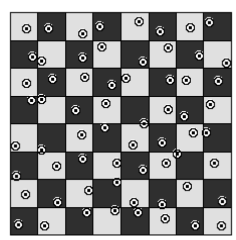
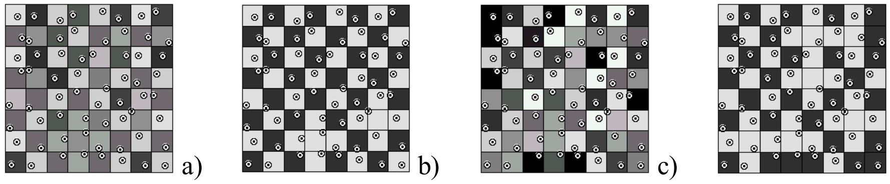


Fig. 3 Sample points & rasterised pattern

If instead we use an interpolation function, which do not honor the observed value *per se*, we get more or less erroneous results (fig. 4).

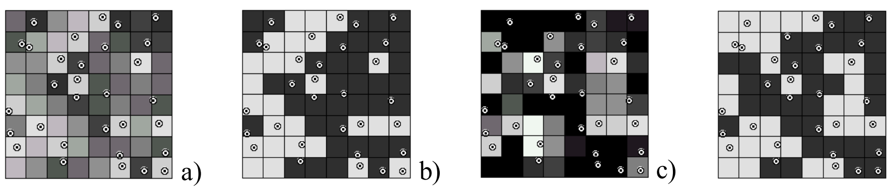
Paper presented at the 3rd AGILE Conference on Geographic Information Science in Helsinki, May 25-27, 2000.



a) b) c) d)

Fig. 4 Interpolated 8x8 raster image from 64 Boolean sample points, randomly placed in each grid cell: a) Inverse Distance Weights (IDW) to 8 neighbors, b) reclassification of (a), c) spline smoothing function to 8 neighbors, and d) reclassification of (c). The reclassification is done as a threshold using the value 0.5. The interpolation methods were set to the default values in ArcView, except for number of neighbors. Both illustrated interpolation methods can be parameterized to get a true chessboard, that, however, demands iterations and skills, as well as a priori knowledge about the pattern of the generated surface.

The great expense of field and inventory work requires the fullest use of existing data and the application of interpolation methods. Hence, the general case is that the sampling grid is much more sparse than the interpolated grid (Fig. 5). Note that the chessboard is somewhere between field and object data – this of course tends to give a bias in the analogy between the landscape and the chessboard. The application of interpolation methods to the irregular samples of the chessboard is rather useless, as also illustrated from a semi-variogram plot of the data (fig.6).



a) b) c) d)

Fig. 5 Interpolated 8x8 raster image from 31 randomly selected points (see fig. 4): a) IDW to 8 neighbors, b) reclassification of (a), c) spline smoothing function to 8 neighbors, and d) reclassification of (c). The reclassification is done as a threshold using the value 0.5.

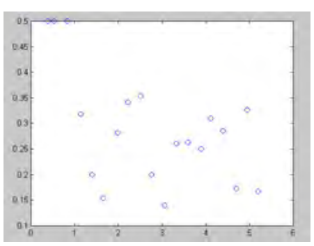


Fig. 6 Semivariogram plot of the data in figure 5. It is clearly no use in applying geostatistical interpolation (kriging) to the data set.

There is a development in spatial interpolation methods for various kinds of data; some of which also can give spatially correlated error estimations (Hutchinson, 1996;

Running and Thornton, 1996). These methods, however, are not part of standard statistical or GIS programs.

Note that the interpolation of the chess-board data is truly 2D, whereas the Earth's surface is a spheroid and interpolation with different geoids and projections render different results (Robeson et al, 1996).

Primary data capture from remotely sensed imagery to GIS is an important part of the integration of GIS and modeling. Remotely sensed data have a definitive grain size and thus resolution. Apart from grain size, problems with sensor quality, spectral properties of the observed phenomena and georeferencing introduce errors when interpreting and classifying remotely sensed data (Fig. 7).

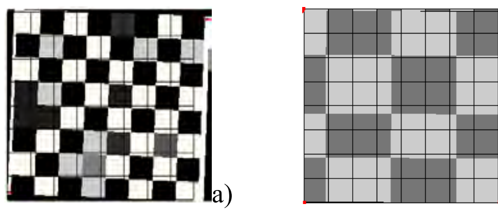

a) b)

Fig. 7 Schematic examples of problems with using remotely sensed data for portraying the earth's surface: a) georeferencing and spectral properties of the observed phenomena, b) grain size and geometrical distortions in the sensor.

3. Monitoring the processes

Having established the playing arena a working hypothesis for the processes occurring need to be formulated. For environmental phenomena an initial inductive approach is almost inevitable. Only after a set of observations is at hand is it possible to use coincidental data to formulate a deductive hypothesis.

Observations of natural phenomena are often done in regular intervals. Satellite images over an area are usually taken at the same time of day, the same goes for many climate station data, water flow and water quality measurements. Measuring the chess game every morning at 09.30 (cheapest because the chess board is outside the coffee room) always gives the same result (Fig. 8).

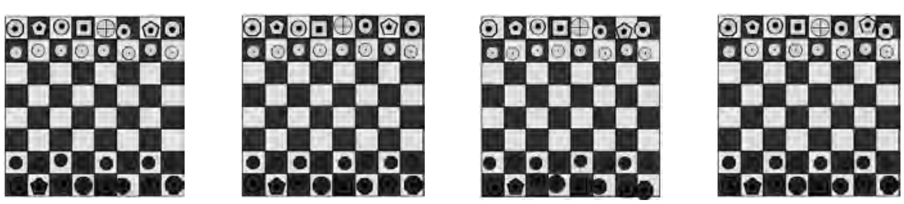


Fig. 8 Observations of a chess game at four occasions. At first the game is apparently static. Only with a more detailed scrutiny it is revealed that the players actually are shifted a little between each observation. However, as we have no hypothesis or information of sub-cell pattern or process we neglect this as observation error.

Our observer works late one evening and by chance observes the chessboard, and *voila*, something has happened. As there is obviously another time scale of the processes than a diurnal (weekly, monthly or annual) our observer starts to observe the game regularly when he works late. A rather erratic series of observations turns up (fig. 9).

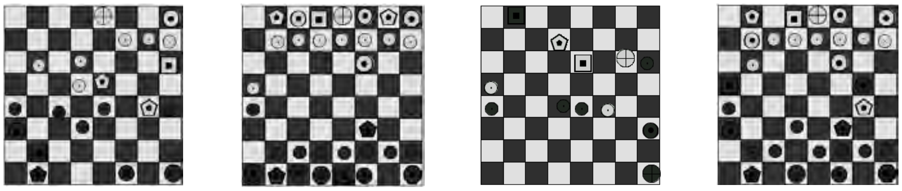


Fig. 9 A series of temporally random observations of the chessboard.

Because of the strange observation angle (from above or nadir in remote sensing jargon) the visualization of the players is poor, and it is difficult to distinguish the actors. However, a few hypotheses on the role of the actors in the game can be put forward:

- one actor, or species, (called bishop) seems to be bound to a certain feature type, or habitat, in the playing ground,
- the smallest and most common actor (called pawn) seems only to be able to move unidirectional like water downhill (however, with a few exceptions – once even reincarnated to a another actor (called queen), and
- the actors in the corners (called rook) seem to be most home bound.

After some months of random observations hypotheses 1 and partly 2 are corroborated whereas 3 is falsified (*sensu*: Popper). After several years of fund seeking the observations can be transformed into intense campaigns during evening time. With observations down to ten minutes intervals some of the rules of the players crystallize themselves, however the role of the one called knight escapes a robust formulation. Finally a sensor connected to a real time observation can capture the full sequence of activities. The role of each player can be formulated.

4. Modeling the full game

The identified role of each player leads to a surge in modeling the game, mostly by using a rule based approach (cf. Fedra, 1993). The formulation of initial and boundary conditions are rather straightforward. The application of an object-oriented approach for each player is favorable; a certain actor can only do a certain action, which cannot be done by another actor.

However, even though the game is spatially defined, it is not possible to use the toolbox of any commercial GIS for playing the game. And only a few softwares have architecture open enough to allow the GIS game to be programmed to them, but with great difficulties (e.g. GRASS). A tight coupling between the game and its arena is thus closed. With a customized GIS it is possible to create a graphical user interface (GUI) that can aid in setting up initial and boundary conditions, and even to allow the set-up of the players positions in the middle of a game, and use that as initial condition. This leads to the development of an intermediate coupling of the chessboard and the game simulator through sharing of a common file format. It is a bit cumbersome to use and never reaches the wide spread use as to improve the societal awareness of the game.

Concerning the game itself the combination of advanced machine learning like artificial neural networks and faster computers, more alternative game outcomes can be foreseen after each activity (draw). Finally one computer (Deep Blue) succeeds in winning the game. This is now more of esoteric interest to the chess community; the general public, policy and decision makers are unaware of the development.

5. Discussions and conclusions

A game of chess always aims at check mate – which is unambiguously defined, as is the role of each player. The role of the game shows no evolution, neither in space, nor over time. If one would change the extent of the arena, the role of the players or the outcome for check mate to an unknown event, the computer would have small chances in winning. In a transient environment that is how the evolutionary game is played.

In the simple case of chess there are only two scales that are of importance, that of a cell and the whole board. Further more, the game as such have no influence on the arena. In a landscape all discretised scales are arbitrarily chose, the real landscape is a continuous nested hierarchy (Allen and Starr, 1982): As shown by Holling (1992) some scales have a dominance generating spatial architectures and temporal cycles, entrapped by key stone species. This also leads to the conclusion that the processes are rather forming the patterns than the other way around – and that the systems has feedback loops at various scales. All those aspects can be disregarded in the special case of the chess game.

The general conclusion that can be drawn is that modeling in GIS is hampered by several shortcomings, that carefulness need to be exercised when using distributed data for modeling, and that the quality of many GIS integrated models is poor; also because they have poor GUIs, fail to visualize the results, and hence do not reach the intended user community.

In order to secure the development of more high quality GIS integrated models the following specific tasks need to be attended:

- closer co-operation between GI model researchers in general, and particular among i) researchers studying the same phenomena but adopting different methods and/or scales, ii) researchers and teachers, and iii) researchers, planners and decision makers,
- sub-grid parameterization, up- and downscaling, and nesting models of different resolution,

- spatial and temporal domain, grain size and sampling intensity when integrating data from various sources,
- strategies for sampling spatial phenomena in order to get representative data,
- development of spatial interpolation methods and spatially correlated error tracking and tagging,
- methods for evaluating the influence of error and error propagation on model performance, and error visualization for communication information on uncertainty,
- integration of remote sensing into GIT models,
- integration of temporal processes into GIT (3D- and 4D-GIS),
- integrated systems that support a complete digital data flow from data collection with mobile field GIS to visualization and exchange of results via networks,
- formulation of versatile criteria for evaluating the prediction power of GIT related environmental models,
- compilation of high quality, accessible (shared) data bases to be used as back-drop for evaluating the predictive power of different GIT related environmental models,
- establishing base line and framework data,
- development of guiding GUIs that can lead the user to select the best method concerning the formulated problem and the available data,
- development of friendly interfaces that promote the dissemination of GIT integrated models to domain experts, planners and managers,
- deeper understanding of semantic issues within environmental models and integrated systems,
- development of a new generation of GIT related environmental models using object orientation and modular approaches and spatial relations, and
- development of generic tools for model formulation directly in the GIS.

References

Allen, T.F.H. and T.B. Starr, 1982. Hierarchy: perspectives for ecological complexity. University of Chicago Press.

Burrough, P.A., R. van Rijn and M. Rikken, 1996. Spatial data quality and error analysis issues: GIS functions and environmental modeling. In M.F. Goodchild, L.T. Steyart, B.O. Parks, C. Johnston, D. Maidment, M. Crane and S. Glendinning (Eds), GIS and environmental modeling: Progress and research issues. GIS World books, Fort Collins, pp. 29-34.

Fedra, C., 1993. GIS and environmental modeling. In: M.F. Goodchild, B.O. Parks and L.T. Steyart (Eds), Environmental modeling with GIS, Oxford University Press, pp. 35-50.

Fotheringham, S., 2000. GIS-based spatial modeling: A step forwards or a step backwards?. In S. Fotheringham and M. Wegener (Eds), Spatial Models and GIS. Taylor and Francis, London, pp. 21-30.

Goodchild, M.F., 1993. Data models and data quality: problems and prospects. In M.F. Goodchild, B.O. Parks and L.T: Steyart (Eds), Environmental modeling with GIS, Oxford University Press, New York, pp. 94-104.

Holling, C.S., 1992. Cross-scale morphology, geometry, and dynamics of ecosystems. Ecol. Monographs, 62: 447-502.

Hutchinson, M.F:, 1996. Thin plate spline interpolation of mean rainfall: getting the temporal statistics correct. In M.F. Goodchild, L.T. Steyart, B.O. Parks, C. Johnston, D. Maidment, M. Crane and S. Glendinning (Eds), GIS and environmental modeling: Progress and research issues. GIS World books, Fort Collins, pp. 85-90.

Maidment, D.R:, 1996. Environmental modeling within GIS. In M.F. Goodchild, L.T. Steyart, B.O. Parks, C. Johnston, D. Maidment, M. Crane and S. Glendinning (Eds), GIS and environmental modeling: Progress and research issues. GIS World books, Fort Collins, pp. 315-323.

Robeson, S.M. and C.J. Willmott, 1996. Spherical spatial interpolation and terrestrial air temperature variability. In M.F. Goodchild, L.T. Steyart, B.O. Parks, C. Johnston, D. Maidment, M. Crane and S. Glendinning (Eds), GIS and environmental modeling: Progress and research issues. GIS World books, Fort Collins, pp. 111-115.

Running, S.W. and P.E: Thornton, 1996. Generating daily surfaces of temperature and precipitation over complex topography. In M.F. Goodchild, L.T. Steyart, B.O. Parks, C. Johnston, D. Maidment, M. Crane and S. Glendinning (Eds), GIS and environmental modeling: Progress and research issues. GIS World books, Fort Collins, pp. 93-98.