Potential and Problems of the Cellular System Approach for Environmental Modeling and Simulation

Jochen Wittmann

HTW Berlin, University of Applied Sciences, Dept. Environmental Informatics, Wilhelminenhofstraße 75A, 12459 Berlin, Germany, e-mail: wittmann@htw-berlin.de

Abstract

By the growing use of geographic information systems (GIS) the use of cellular approaches for modelling and simulation spatial and especially geographical processes and systems gains more and more importance. This paper gives an overview over the basic definitions and emphasizes a proper differentiation between the very tight definition of real cellular automata and the very universal one for cellular systems.

On the base of these definitions typical problems are discussed, which raise for modelling spatial objects given by raster or vector data and by representing spatial relationships between these objects. Furthermore the difference in semantics between neighbourhood relations in GIS and neighbourhood relations needed for modelling purposes are elaborated.

1. Cellular Automata and Cellular Systems

For a long number of years cellular automata serve as a paradigm for modelling real world systems. In the late 60th J.H.Conway developed the automaton “Life” to model processes such as cell nascence and death caused by aggregation and isolation. Later on a theoretical foundation for the model class of cellular automata has been developed (e.g. Banks [1] and Wolfram [5]). For the application area of environmental modelling these approaches recently re-gained actuality. Because of their acquaintance to the raster data offered by geographic information systems (GIS) a large number of models dealing with geographical processes are developed on the base of cellular automata. A representative selection of examples for this type of models is found in the anthology edited by Goodchild [2].

However, within these examples and most of the application models found elsewhere in literature, the term cellular automaton is used for a very broad range of modelling approaches which suggests an engagement with the goal of more precise differentiations and definitions.

But the following definitions do not only serve theoretical purposes. For the practitioner the mathematical formalism this paper gives in the following section opens the range of successively more detailed model specification between the rigid and close definition of cellular automata themselves and the very general definition of a so-called cellular system.

On the one hand, giving these definitions the modeller will obtain a feeling for potential problem fields in transferring real system dynamics to cellular systems and on the other hand he gets a guideline how to use the mathematical concepts to model system behaviour more detailed in a formally proper way.

Doing so, the next section gives two definitions, which will be our base for a discussion of practical modelling problems later on.
1.1 Definitions

A classical definition for a cellular automaton is given by Wunsch in [6]:

A **cellular automaton** is a 5-tupel

\[ \sum = (Z, Z, R^m, f, \lambda) \]

with:

1. \( Z \) is a set of states, with elements coming from the state alphabet \( Z \).
2. \( Z \) is a non-empty set with \( |Z| \leq |N| \), with \( N \) as the set of natural numbers.
3. \( R^m \) is the spatial extension of the automaton with \( R^m \subseteq G^m \), \( m = 1, 2, 3, \ldots \) and \( G \) the set of whole numbers.
4. \( f \) is called the state transition function with \( f : \{R^m \to Z\} \to Z \).
5. The neighbourhood relation \( \lambda \) lists for every space element a set of space elements which are neighbours to the given one:
   \( \lambda : R^m \to \wp(R^m) \).

This definition describes an equidistant raster on the (geographical) space \( R^m \) with equal cells. In each of them works a state automaton with an individual internal state and a state transition function. All the cells work with the identical state transition function. The states of the cells change discrete in time and value going from state \( k \) to state \( k+1 \). The states of the neighbours of a cell may influence the state transition.

To apply this concept for real world problems in environmental modelling the following problems will arise:

1. The discrete alphabet given by the definition opposes to the need of real value variables for system quantities.
2. Each cell can hold only one state variable. A typical geographic space unit is characterized by a set of parameters, however.
3. The automaton only models a sequence of states without any further relation to the time axis.
4. The cells do not have any outputs and inputs in contrast to input- and output functions given by system theoretical concepts.

Caused by these restrictions a more general definition for a so-called cellular system is given, which meets all the requirements of spatial modelling. Again we define in accordance to Wunsch ([6]):

A **cellular system** is a 6-tupel.
\[ \sum = (T, Z, R^m, \lambda, f, \leq) \]

with:

1. time base \( T \) is a non empty set with the relation \( \leq \) defining an order on its elements.

2. \( Z \) is a freely definable set called the state set for the system.

3. \( R^m \) is the spatial extension with \( R^m \subseteq G^m, m=1,2,3, \ldots \) and \( G \) being the set of whole numbers.

4. \( f \) is the state transition function with:
   \[ f : \{R^m \to Z\} \times T \times T \to Z \]
   It transfers a state \( z \) at a space element \( r \in R^m \) and at a point in time \( t \) to a state \( z' \) of the same state element at time \( t' \) (\( t \leq t' \)).

5. The neighbourhood relation \( \lambda \) lists for every space element a set of space elements which are neighbours to the given one:
   \( \lambda : R^m \to \wp(R^m) \).

In comparison to the cellular automaton we notice the following extensions:

- By the additional time set and its ordering relation time-related processes as needed for system simulation can be mapped.

- All the model variables used in the definition are interpretable as vectors. Doing so, the restriction of only one state variable per cell is revoked.

- The variables are defined without any relevant restriction and therefore can be used for universal modelling purposes.

- The neighborhood relation is unrestricted as well.

- There is no advice concerning the spatial collocation of the cells.

From these observations we can state that this definition gives really a non-restrictive formal frame for modelling. On the other hand, however, it gives only few constructive advise how to build a (spatial) model. The modeller gets all freedom to fill the frame individually, but just this range of freedom leads to conceptual problems which shall be discussed in the following sections.

2. Basic problems in modelling with cellular systems

By practical reasons, the semantics of space used in modelling environmental systems is mostly pre-specified by the concepts given by geographic information systems. The simple cause is that all the data to
use within the model will come from there. So we first will have a look on the means to model spatial sys-
tems and relationships given by recently used GIS.

Geographic information systems provide two very different basic concepts for describing spatial ob-
jects: raster data and vector data.

2.1 Modelling with raster data

In the case of raster data a free choosable but equidistant raster is put over a geographic region. It is obvi-
ous to identify each of these raster cells with one of the cells of a cellular automaton.

So the advantages of this representation are perspicuously:

- easy to handle
- GIS data are easily transformed into model data
- Free scalable model accuracy by free scalable raster width.

In addition to these arguments the approach is easily adaptable to higher dimensions of space by going
from squares to cubes, from cubes to hypercubes and so on.

Problems using raster data are:

The raster has to be equidistant and perpendicular and the width of the raster determines the scale of the
model.

In the first case the problem lies in the effect of discretisation if an object lies cross to the raster. Caused
by this argument the modeller will be forced to work with a quite small raster width to reach sufficient ac-
curacy.

But doing so, another problem arises: There will be some real world objects to model that spatially extend
over a set of cells. Figure 1 gives an impression of the situation for the modeller between raster representa-
tion and vectorized objects.

![Fig. 1: values of model quantities in raster and vector representation](image-url)
For those real world objects, which extend over a set of raster cells, a transformation function is needed: Firstly to summarize the values for this spatial object from the attribute values of the cells covered by the object. This has to be done by a separate aggregation function which is not mentioned in the standard-definition of a spatial system.

The analogous problem has to be solved for the other direction: An attribute value of a spatial object has to be mapped on the set of cells the spatial objects consists of. Not in all the cases a simple homogenous distribution models the real world circumstances correctly.

A typical example for these problems are population distribution over a geographical and administrative region such as a municipality: To work with statistical data the extension of the spatial unit has to be the community itself, the distribution of the people within this region will not be homogenous at all and will be one of the main interests of the model.

Let us transform this situation to raster GIS in which the municipality is represented by a set of raster cells:

1. First direction, the aggregation function: The over-all population is the sum of the population of all the cells.

2. Second direction, the distribution function: How does the population distribute over the cells representing the municipality?

Both functions have to be modelled in addition to the definitions of cellular systems and/or cellular automata!

As these deliberations show, in this modelling task lies the complete scaling problem which is very well known from the geographic modelling part. The aggregation functions and the distribution functions are not trivial at all (for a more detailed discussion of this range of problems see Ortmann in [3] or Wittmann in [4]) and lead to hard complications in contrary to the seemingly very simple modelling approach by cellular systems.

### 2.2 Modelling with vector data

The alternative is given by a vector-representation of geographical objects as used for vector-GIS. It is obvious that the discretisation problems will completely disappear using this approach. The modelling accuracy is determined by the accuracy of the defining vector line. An aggregation function and/or a distribution function between the different scales of spatial object and raster cells is no longer necessary.

However, this modelling approach suffers under the free editable border line of a spatial model object. Without any regularity given by a raster, these borders are freely editable now. The consequences are:

1. Regions of different sizes:
   The values for all the variables used in the model have to be normalized before brought into relation to values of any other spatial model object.

2. Very complicated neighbourhood relations because of the irregular size of the objects: The number of neighbours of a cell will be variable which causes a much more complicated specification in the model description part. The necessary information strongly exceeds the simple definition of cellular automata.
This first argument requires the attention of the modeller but can be solved with reasonable effort. The second argument shall be discussed in detail by the following section.

2.3 Neighbourhoods

For cellular automata and for the corresponding raster-GIS representation the selection of a suited neighbourhood relation is normally restricted to a selection among the classical relations (von Neumann, Moore, …). It should be remarked here that the modeller has much more freedom in modelling concerning the neighbourhood relation, however. Any relation that can be expressed by a filter or a window given by a set of cell indices in relation to the index of the cell under observation is allowed by the definition. This can be done very simply by the regularity of the raster.

For cellular systems and vector GIS the neighbourhood specification will change substantially: There is a completely irregular mosaic over the geographical space given by free definable border lines of the spatial objects. The neighbourhood is no longer defined by indices of raster cells but by an individual set of neighbours for each cell.

A typical formal (and mathematical) representation of the neighbourhood relation for all the spatial objects used within a model would be a neighbourhood matrix as shown exemplarily in figure 2.

![Neighbourhood Matrix](image)

Fig. 2: An example for a neighbourhood matrix for a spatial area with six sub-regions

Beneath these implementational remarks using vector representation substantial conceptual problems will arise at two points:

**Geographical versus logical neighbourhood**

All neighbourhood relation given by cellular automata has to be interpreted as a geographical neighbourhood. This might result in some discrepancies for modelling. One could imagine that two cells are near neighbours concerning a Moore-neighbourhood but concerning the semantical relation needed for the model they are completely isolated to each other.

To give an example: two rural regions are spatial neighbours but separated by a motorway. Therefore the logical neighbourhood which models reachability is completely different to the one given by the geographical position of the regions under observation. Because of the regularity of the cellular automaton, this situation cannot be modelled by this concept. Only one neighbourhood relation is allowed. And this neighbourhood is given by the spatial adjacency of the cells.
Using cellular systems this situation could be represented by an individual mark in the neighbourhood matrix. The matrix could be set by the logical neighbourhood relation needed for the modelling purpose. Again this concept is open for any interpretation (logical neighbours) but requires additional information (specification of the semantical neighbourhood) on the other hand.

**Differentiating logical neighbourhoods**

The second problem deals with the differentiation of the neighbourhood relation for the different components of the state vector of a cell in space. Because the vectorization is only allowed for cellular systems, this problem will appear for this more extended concept only.

To give an example for the problem: A model for the settlement of a population over a geographical region will refer to the attributes “shopping facilities” and “leisure facilities” for a given cell. The facilities are determined in relation to the existence of shopping centres and recreational facilities in the neighbourhood of the cell under observation. It is obvious that the inhabitant of a cell feels the shopping neighbourhood completely different to the leisure neighbourhood: People will accept a quite long car ride to practice leisure activities but they like the comfort buying their bread rolls in direct neighbourhood to their residence. This differentiation cannot be expressed by cellular automata at all.

The mightiness of cellular systems would allow a mathematical representation of this situation: For each of the attributes given in the state vector a differentiated neighbourhood relation can be given. The mathematical apparatus allows this by specifying a vector of neighbourhood matrixes.

For practical reasons the expenditure in defining all these neighbourhoods has to be taken into account, however. Although GIS functionality can support this task technically, much deliberation is necessary for a reasonable specification of the attribute-dependent neighbourhood relationship between the cells. This should a modeller take into account considering the cost–value ratio for his model.

It should be pointed out that there is a fundamental difference between modelling in GIS and systems modelling with dynamical processes. GIS administers the geographical relations of spatial objects. Simulation models extend this specification to logical attributes and relations, which should be stored separately. Under this point of view the geographical neighbourhood represents just one of the open set of “logical” or “semantical” neighbourhoods needed for system modelling.

### 3. Conclusion

For the practitioner in modelling there are two general ways to build a model with geographical related objects using cellular approaches:

First would be the “easy-to-do-approach” by cellular automata. This formal concept gives a very strict frame for the modelling task and can easily be handled algorithmically. However, for most of the applications its formal expressiveness does not fulfill the needs of the modeller.

So for real world modelling projects more or less extensions are necessary. These extensions can be summarized under the formal concept of cellular systems. All restrictions are revoked by this definition. The modelling task however gets much more complicated by the design decisions the open concept of cellular systems demands. Taking into account the deliberations about the representation of spatial objects and the neighbourhood relations the modeller has to decide the following two points of interest:

1. Homogeneous cells versus individually sized cells.
2. Homogeneous neighbourhood for all the components of the state vector versus attribute specific “logical” neighbourhoods for each state variable representing the situation of a cell.
In dependence on the decision between the free specifiable cellular system and the restrictive cellular automaton the expense for implementation of the simulation algorithm and for setting a consistent initial state for the model has to be considered.

This paper has the intention to give an overview over the range of modelling spatial systems using cellular formalisms. Main intention was

- to give criteria for proper use of the terms cellular automaton in contrary to cellular system and
- to give hints for a developer of such models by showing on the one hand the options in modelling and on the other hand the respective decisions will cause.
- To give an impression about the dangers when GIS data structures are used for modelling purposes without thinking about the consequences on semantics in model specification.

Especially for modelling and simulation environmental models the concepts given by cellular systems form a proper formal and mathematical frame, which should be used to approach spatial modelling problems. However, nothing is as easy as it seems in the beginning: Even using the easy understandable concept of cellular systems, the necessary extensions for modelling real world problems lead to complex and highly sophisticated model specifications!

4. Literature