A Theory of Spatial Reference Modes and System Archetypes

(Presented at System Dynamics Winter Conference in January 2011, Austin Texas)

Todd BenDor* and Nikhil Kaza**

Department of City and Regional Planning

University of North Carolina - Chapel Hill

New East Building - CB# 3140

Chapel Hill, NC 27599-3140 (USA)

* Corresponding Author, Email: bendor@email.unc.edu, Ph: 001-919-962-4760

Draft Copy: Please do not cite or distribute. Feedback would be warmly appreciated! –

***bendor@unc.edu, **nkaza@unc.edu**

Abstract

The development of historic modes of dynamic behavior is widely accepted as a key step in the system dynamics modeling process. By understanding past system trajectories, modelers can delineate causal relationships within dynamic systems, particularly by employing dynamic system archetypes such as growth and decline (exponential, goal-seeking, etc.), oscillation, and combinations thereof. Our goal is to characterize spatial dynamic patterns in a similar manner to the current characterization of non-spatial dynamic system archetypes. We extend the reference mode concept to models of spatial-dynamic phenomena, focusing on archetypes of changing spatial patterns in multi-dimensional landscapes using two characterizations of space, fields and 'networks'. While fields, as they are known in spatial science parlance, provide a continuous description of space, we argue that networks more readily characterize the discretization of space.

Recent spatial-system dynamics research has articulated 'space' as a tessellation into regular grids. Similar tessellation can be employ hexagons, triangles, and other geometric shapes. Although this is quite common in the geography and spatial modeling literature, there is often little underlying logic that guides decisions on the representations of space in these models. We argue that in order to abstract away the artifacts of this tessellation, we should instead view spatial interactions as they occur across a topological network that defines the underlying structure of space. By doing so, we can construct and use irregular tessellations of space and then accommodate diverse spatial representations, including raster and vector models of landscapes, social connections and networks, and diffusion vectors.

In this paper, we explore the connections between temporal dynamics and their spatial manifestations of change. We tap a growing literature on static spatial analysis techniques and spatial network representations to better understand the influence of space on dynamic relationships. We also explore several factors in creating spatial-dynamic archetypes, including the expression of particular growth and collapse patterns, and the spatial contiguity necessary for temporal and spatial feedback. In particular, we apply these ideas to a variety of spatial problems including urban growth, ecological systems, and networks (disease transmission).

By extending the reference mode concept spatially, we argue for a spatial modeling paradigm that parallels the "learn-by-analogy" pedagogical technique presented by system archetypes that have evolved during the last fifty years of system dynamics research.

Introduction

As systems dynamics (SD) has growth in popularity and range of application over the last fifty years, its use of scientifically rigorous and iterative modeling processes has differentiated it from other modeling methods (Saeed 1998a; Saeed 1998b; Saeed 2001). A series of efforts have been made to explicitly structure the SD modeling process (Sterman 2000). In particular, application of historic modes of dynamic behavior, known as "reference modes," has become a key factor in promoting SD models that are rigorous and causally-focused (Saeed 1992; Saeed 1998a).

Reference modes are storehouses of sorts for dynamic information, allowing modelers to explore historical dynamic patterns of systems to better understand how systems behave over time. This information is used to create a causally-explicit, dynamic hypothesis of how a system operates and how problems may develop, which can then form the basis of rigorous, quantitative stock-flow-feedback representation of system elements (Sterman 2000). As SD modeling has become more common, modelers began creating archetypical dynamic hypotheses known to produce frequently encountered reference mode behaviors. These 'systemic archetypes,' (sometimes also referred to as 'generic structures,' 'atoms of structure,' or 'micro-structures,' which different authors define and use differently; Paich 1985; Lane 1998; Wolstenholme 2003; 2004), help to explain a variety of system behaviors, the most basic of which include linear, exponential, and logistic growth and decline, oscillations, and overshoot and collapse (Breierova 1997 ; Chung 2001). Wolstenholme (2003, pg. 342) makes an excellent case for the development and use of archetypes, noting their ability to "offer solutions to complex problems,…aid quantitative modeling, ….assist model conceptualization, …[and] communicate modeling insights by collapsing a model down to its basic loops."

While there have been several applications directly within the field, a vibrant field of spatial-dynamic modeling has emerged in the last two decades outside of SD, offering compelling arguments for explicitly considering detailed spatial structure and effects within models. In fact, work on spatial autocorrelation has demonstrated major specification errors and other problems in models that fail to explicitly consider the spatial relationships between interconnected system elements (Anselin 2002). For example, in multi-species ecological, populations can be modeled as they change at different rates. However, when interactions between species are crucially dependent on their locations, not just on aggregate numbers, it pays to make the spatial dimensions of these populations and interactions explicit.

Unfortunately, very limited work has attempted to apply the rigorous elements of the SD methodology in a spatial context, particularly using well-developed spatial analytical frameworks advanced in recent decades. The application of reference modes and systemic archetypes in the spatial realm is very much a new frontier for SD research, with substantial implications for the rigor and communicability of spatial-dynamic models.

During this article, it is our goal to offer a theory and strategy that extends system archetype concepts to dynamic systems whose structure and behavior are determined by spatially explicit processes. In exploring this extension, we focus on expanding current two-dimensional reference modes (point data mapped through time) into four/five-dimensional modes (point data mapped over a two- or three-dimensional spatial surface and through time).

This article is organized into five substantive sections, beginning with a comparative discussion of spatial reasoning in SD and other fields, followed by a discussion of temporal and spatial feedback and a taxonomy of continuous spatial-dynamic processes. We then offer several examples of spatial-dynamic models, including simple spatial extensions of basic system archetypes (which we term 'extensive processes'), followed by more complex, 'intensive process,' examples of spatial diffusion, simple disease spread, and disease spread across a dynamic spatial network. Finally, we conclude with a discussion of the implications of this research on the larger system dynamics research agenda.

Space in System Dynamics

System dynamics has explored spatial modeling a number of times over the last 50 years. Zonal models, such as the one created by Wilbert Wils (1974) to extend the Forrester (1969) Urban Dynamics study, have attempted to disaggregate areas, such as cities, by replicating model structures to represent varying characteristics of the landscape (e.g. central business district, inner ring suburbs, outer ring exurban areas). However, although Urban Dynamics offered sophisticated dynamic representations of urban development processes (even in today's terms, more than 40 years later), representation of spatial heterogeneity was so limited as to amount to a major criticism of the model and its later extensions (Burdekin 1979).

More recent zonal models include work by Mosekilde et al. (1988) who model chaotic behavior in a two-zoned city, Rich (2008) who modeled the movement of foot and mouth disease between zones throughout South America (Figure 1a), and Pfaffenbichler et al. (2010) who model land use-transportation interactions in the City of Leeds, UK. These studies are similar in their attempts to spatially-disaggregate the area of analysis in order to more accurately parameterize models, understand interactions, and improve model usability and accuracy.

[Insert Figure 1 here]

The problem in confining SD spatial reasoning in this manner relates to the manner in which zonal models treat space. For example, in the Wils (1974) model, we may know, for example, that Zone 2 lies between Zones 1 and 3, and that it possesses some spatial extent

(Figure 1b. However, within the model itself, that extent is irrelevant, and zones are modeled as two interacting entities without any specific location. Zones, like aggregate models, continue to represent spatial areas as points, which fail to convey any information about relationships across or within space, or information about spaces themselves. Although this representation may be sufficient in many situations, it is limiting in many others, particularly scenarios where substantial environmental or spatial heterogeneity determines or influences system structure and behavior (e.g. Anselin 2002; BenDor and Metcalf 2006). As Douglass Lee (1973) discussed in his seminal “Requiem for Large-scale Models,” much of the usefulness in modeling arises when models are used to represent sophisticated problems in usable ways. For many problems, spatial detail greatly enhances model accuracy, visualization and communication ability (Lowry and Taylor 2009), and usability.

More advanced spatial applications in SD include ’s (1999) simple SD model of spatial heterogeneity in a drainage basin, which employed a more sophisticated characterization of gridded space whereby single stocks represented water levels in connected landscape areas (Figure 1c). Ford ’s (2009) model demonstrates the difficulty in replicating system dynamics models in each grid cell, similar to zonal applications. Efforts to overcome this difficulty have emerged in several efforts to spatialize system dynamics models (Maxwell and Costanza 1997b; Ahmad and Simonovic 2004).

Perhaps the most sophisticated effort to explicitly marry SD techniques to spatial modeling have emerged in systems such as the Spatial Modeling Environment (SME), a platform for ‘spatializing’ system dynamics models by replicating them into gridded cells (see Figure 1d) and parameterizing them with geographic information systems (GIS) spatial data (Maxwell and Costanza 1997a; b). However, while this, and similar frameworks, are useful for a variety of applications (Voinov et al. 1999; BenDor and Metcalf 2006), none of the efforts to spatialize SD modeling have attempted to 'spatialize' SD's actual modeling process or its theoretical and scientific underpinnings.

Spatial Thinking in other Disciplines

The field of spatial analysis has growth rapidly in parallel to the development of system dynamics, drawing an array of spatial analytical techniques from fields such as ecology (e.g. tools for assessing the spatial fragmentation of wildlife habitat; McGarigal and Marks 1995) and economics (e.g. spatial econometrics; Anselin 2002; 2003).

Allen and Hoekstra (1993) propose an interesting allegory for spatializing SD theory in their discussion of the 'grain' and 'extent' of ecosystems and ecological communities. In SD, modelers typically focus on determining time step and time horizon, two measures of the 'grain' (temporal resolution, in this case) and 'extent' (length of model run) of a system being modeled. In considering grain and extent in a spatial context, we must consider that behavioral reference modes are empirically observed phenomena and are therefore vulnerable to changes in the scale of analysis (the spatial extent we model) and the unit of analysis (the grain or resolution of space we consider; Wolfram 1983; Allen and Hoekstra 1993). The role of, and sensitivity to changes in, spatial extent and resolution is a profoundly important and on-going area of study in spatial analysis and modeling fields.

An attendant debate within Geographic Information Science (GIScience; Longley et al. 2005) is the conception of space as either Newtonian or Leibnitzian (Galton 2001). The Newtonian conception requires the underlying geography to be absolute and act as an inert container; objects acquire properties, such as position, velocity etc., within this geography. Newtonian space is specified independently and prior to the description of objects that inhabit it and is therefore an absolute view of space. Contrasting this is the more relativist Leibnitzian model, which asserts that space is constructed through relations between arrangements of objects. Therefore, space does not exist in any absolute way, and is merely a construct generated from the locational attributes of our objects of interest. While both views have different merits and problems, we argue that, for the purposes of this article, the Newtonian conception is more readily amenable for use in SD modeling practices (although this may not be true for many of the emerging SD applications in agent-based modeling; Pourdehnad et al. 2002; Borshchev and Filippov 2004). Although it is important to understand different theoretical representations of 'space', we are much more interested in the topological construction of space itself.

While space has been defined in a variety of ways, the spatial science literature has focused primarily viewed space through vector or raster frameworks. In vectorized space, objects are depicted as points, lines (connected points), and polygons (area enclosed by connected lines). In rasterized space, which is more common for spatial modeling applications, space is tessellated into a collection of plane shapes with no overlaps or gaps (sometimes squares, rectangles, or hexagons of equal shape and size, as in a grid). However, as we will argue shortly, the raster-vector debate found in the GIScience literature becomes somewhat irrelevant for our purposes if we are primarily concerned with the topological connectivity between interacting entities in order to define tessellations or vector arrangements of objects.

The vector/raster comparison is similar to that of continuous and discretized models of time in classical SD modeling treatments. While the vector representation of space is more accurate (as is a continuous representation of dynamics), it is often computationally and theoretically intractable for modeling applications. Conversely, raster representations, like discretized time steps, approximate spatial processes given the spatial resolution of a model.

The technical representation of raster and vectors is the manifestation of an important dichotomy underlying the conceptualization of space. The geographic modeling literature (Couclelis 1992; Goodchild 1992; Egenhofer et al. 1999) characterizes this dichotomy by distinguishing ‘fields and ‘objects.’ Field-based representations of space completely and exhaustively tessellate space either into rectangular or other polygonal entities. Once a tessellation is specified (e.g. a rectangular grid or zones comprising cities or suburban regions), each location is endowed with continuous (e.g. temperature) or discrete (e.g. population) attributes, which are subject to change over time due to influence of the attributes of neighboring cells.

On the other hand, objects are entities with attributes that can include location. Therefore, objects can potentially move in space and acquire new attributes. Couclelis (1992) argues that both fields and objects are representative of various types of geographical knowledge and neither uniquely or completely fit the types of problems that spatial system dynamics models may seek to address. The object/field dichotomy is important to distinguish when constructing models that have objects that change locations or locations that have attributes. For the most part, SD models deal with the latter, even when the underlying space need not be exhaustively and continuously tessellated.

The growth of spatial statistics has spawned a new analytical perception of space, which replaces information about the actual location of objects with a network representation of their relationships to each other. These “network topologies,” as they are known, can be powerful representations of space, and can include information about the neighborhood around objects. In addition to representing the topology of the space, this representation lends itself to representing the strength of network relationships through the imposition of weights on network links. (e.g.

strong social relationships, or speed limits determining rate of movement between cities), and abstract information, which may be vitally important to studying a system, about space itself that often cannot be captured by spatial grids (e.g. a disease spreading across a series of valleys, or a flow of information from one local economy to another nearby).

Defining relationships in spatial dynamic systems commonly relies on measures of distance in a landscape or between system elements. Distance is often measured as simple proximity, but under network characterizations, distance can also be modeled in a more sophisticated manner through the use of 'spatial weights matrices' (Anselin 2003), which are arrays that define 'adjacency' in space, or reduce the bulk of information about spatial arrangement in a landscape to a simple representation of neighboring relationships (and strength of relationship) between landscape elements.

System dynamics research has made several forays into network analysis. Reggiani and Nijkamp (1995) explored complexity and chaos across a network, demonstrating "how a network can be conceived of as a complex space-time system, whose evolution depends on critical factors that are interrelated in space and time by means of a connectivity structure." This important finding has led to more complex views of networks, such as that of Cruz and Olaya (2008), who created a network model using the *Mathematica* software that simulates network marketing as it could occur dynamically across changing connections within a network. Additionally, Alekseeva and Kirzhner et al (1994) discuss material exchange across a network, implementing a complex, multi-stock model of a multi-centric, immuno-dependent tumor.

Spatial Reference Modes and Systemic Archetypes

Spatial modes of historical reference behavior represent descriptive patterns of spatial change over time. Usually, these are based on historical observation, and rely, like classical reference modes, on pattern recognition to understand the type of dynamic observed. Like their non-spatial counterparts, spatial reference modes are empirically observed phenomena that are vulnerable to changes in the scale, extent, and grain of analysis (Chen and Pontius In Press). Following this, spatial systemic archetypes are theoretical abstractions that describe, in part or in whole, one or more reference modes.

Taxonomy of Continuous Spatial-Dynamic Processes

Classic SD texts such as Sterman (2000), Paich (1985), Lane (1998) and Ford (2009) define numerous system archetypes, including those grouped around growth, decline and more complex combinations of simpler archetypes (see Table 1). Growth archetypes include linear, exponential, goal-seeking, and logistic growth – a combination of exponential and goal-seeking growth under shifting feedback loop dominance (Glick and Duhon 2001). Similar archetypes describe decline behavior, including linear, exponential, and goal seeking decline. Archetypes can be grouped into more complex archetypes, such as oscillation, damped oscillation, and overshoot and collapse behavior, as a means of avoiding wasted model-building effort and enhancing transferability of basic modeling concepts (Paich 1985; Wolstenholme 2003). Very advanced system archetypes have been developed over the years to represent complex behavior that is commonly seen in many situations (e.g. market growth, Forrester 1968; acceptance-rejection behavior; Ulli-Beer et al. 2010).

[Insert Table 1**]**

We divide spatial-dynamic processes into two different categories, extensive and intensive, in order to explore ways of thinking about systemic archetype structures underlying different types of continuous spatial-dynamic behavior.[1] Extensive spatial processes involve

[1] System dynamics models classically do not consider discontinuous processes (e.g. discrete event modeling; Banks

change at the margins (i.e. processes that flip a point in space from being within a domain of the process to outside the domain, or vice versa). Under Intensive spatial processes, on the other hand, the value of the process at each point affects the value at the neighbors. While, it may seem that extensive processes are a subset of the intensive processes, it will be useful to think of them separately in formulating spatial system archetypes.

Extensive Processes

Extensive processes describe the extent of boundaries or characterize changes in boundaries over time. An example of this would be a model of the region into which a given technology has diffused, with the edges of the region gradually changing as new areas adopt the technology. Extensive spatial processes are analogous to Markov processes, where the value at the next time step is dependent only on the current value, not on history (Bhat and Miller 2002). In a sense, these are strictly binary processes, where Newtonian space is divided into regions either inside or outside the domain of the process.

Under these conditions, the process of expansion of the boundary can be described by archetypes that are very similar to those of aspatial process. Table 2 provides a series of example analogues to the basic system archetypes discussed in Table 1. In this case, rather than describing changes to a bank account, or population, the equations describe changes in the area of a circle, as expressed in changes to the radius of that circle. We visualize this extensive process as changes in the extent of the circle over time, which we depict in the final column as a series of nested circles that depict the circle's growth through time (numbers in each graph show connections to the aspatial archetypical behavior over time).

et al. 2004). Under the same line of thinking, we also do not consider discrete event modeling (such as spatial Poisson processes; Cox and Isham 1980) in the spatial context.

While BenDor and Metcalf 's (2006) study of the spread of an invasive insect (Figure 1e) was an inherently intensive process (described below; an aging chain model determined species density in each 30m grid cell), output was primarily assessed based on the extent of species spread. Creating rules for spread dynamics involved specifying the neighborhood into which insects could travel. This neighborhood was partly dependent on the size of the dynamic time step chosen. Choice of a large time step would necessitate a larger spread neighborhood, or else spread would be artificially slowed, as insects would be technically unable to move great distances in successive time steps.

This issue can be seen in Figure 1e, which depicts varying possibilities in Voinov et al. 's (2007) Patuxent landscape model, which models water flow between surrounding grid cells, where neighborhoods consist of a) contiguous cells only, b) a larger, second ring of cells, and c) a dynamic structure where distance of flow from a cell is based on water depth. This example illustrates the complexities of linking neighborhood size, structure, and dynamics to time step and dynamic processes modeled.

[Insert Table 2**]**

Intensive Processes

Continuous spatial processes can most easily be characterized by graph theory, the mathematical underpinning of network theory (Diestel 2006). A graph G is a collection of set of vertices V and edges E, which define topological relationships between vertices. For a given vertex v, $N(v)$ is the set of all neighbors of v. It is now sufficient to re-characterize space as a tessellation, where the polygons are represented by vertices and the topological connections as edges (see Figure 2).

[Insert Figure 2**]**

Many different kinds of underlying spatial entities can be represented using a network. Figure 2(a) is representation of contiguous polygons that affect one another through their neighborhood spillovers. Similarly a regular grid translates to a near regular graph (Figure 2b).[2]

Figure 2(c) on the other hand is representation of non-contiguous polygons. However, the processes in one of these polygons may affect its nearest neighbors, irrespective of whether those neighbors share a boundary.[3] It is therefore, important to realize that contiguity does not guarantee connectivity. Rather, connectivity is determined by the problem in question and the particular spillover effects that necessitate modeling. For example, 'second order' contiguity, a measure commonly used in spatial statistics, can be represented in a simple graph even though it necessitates links between polygons that are typically one link removed from each other (i.e. imagine a neighborhood constructed entirely out of your neighbors' neighbors; see Figure 2d and note that neighboring polygons are not connected in the network).

Once such network is constructed, the archetypical patterns are fairly straightforward to construct, and we can characterize the behavior of any given node S_i as a function of its own dynamics, and the dynamics of its neighboring nodes $S_{N(i)}$, respectively.

$$\frac{dS_i}{dt} = f(S_i, S_{N(i)})$$

Temporal and Spatial Feedback

In George Richardson 's (1999) landmark work on feedback theory, he proposes that in modeling dynamic systems, the direct or indirect influence of a system element on itself is based on contiguous temporal relationships. Spatial analysts often assign causal relationships to spatial

[2] A graph is said to be 'regular' when the 'degree' of all vertices (defined as the size of a given vertex's neighborhood) is equal.

[3] Conversely, it may be possible that the space could be represented as clusters of disconnected components (e.g. isolated areas), rather than a connected graph. However, this does not affect the construction of archetypes since the processes in each isolated component do not affect each other, allowing us to model processes in each component independently.

behavior, but this is not possible without time. Spatial 'causality' does not exist; time mediates spatial relationships, determining whether one object, affecting another across space, forms a causal influence with respect to time. A change in a certain grid cell, for example, can only affect other, surrounding grid cells, later in time. This means that the uni-directional causal perception in SD, which models time as an arrow moving in one direction, becomes more complex when time establishes causal relationships that form patterns across space. Using this logic, we can see the potential problems in transferring ideas of causality from time to space. This concept is fundamental to understanding feedback that occurs through space.

Since time relentlessly marches forward; the past can only influence the future and not vice versa. We can consider spatial feedback to be "bidirectional," in the sense that neighborhood relationships are more often than not, bidirectional relationships. Unidirectional topological relationships are certainly possible and are useful in some cases, such as flow from higher elevation to lower elevation, and one-way streets (network representations allow for directed networks to be constructed). However, predominantly undirected networks represent the topological relationships between spaces, and processes at one point (or node, or cell) not only influence all of its neighbors in the next time step, but simultaneously all the neighbors influence the process at that point in that time step.

It is therefore important to differentiate between concurrent dynamics and sequential dynamics; that is, determining how fast given dynamic processes occur versus how fast those processes influence surrounding neighbors (e.g. spread or diffusion). Furthermore, because the SD models are constructed on a 'serial' computer, it is imperative to understand the quirks of software in handling concurrency (e.g. software can number cells/nodes/points and calculate dynamics in each sequentially, or it can move North to South and West to East calculating in order of cell/node/point position).

Examples

We can now characterize any of the basic spatial system archetypes listed in Table 2 using arbitrary graph structures to characterize tessellations of space. To do this, we create a random array of nodes, connected through a random graph using *Netlogo* 4.1, a spatial, dynamic, and agent-based modeling framework (Wilensky 1999). A number of software tools now exist for performing network analysis, including *NetLogo* (developed at Northwestern University), *AnyLogic* (developed at XJ Technologies in St. Petersburg, Russia; Borshchev and Filippov 2004; http://www.xjtek.com/), *SWARM* (originally developed at the Santa Fe Institute; http://www.swarm.org/), and the *Recursive Porous Agent Simulation Toolkit* (*REPAST*, originally created at the University of Chicago; Collier and North In Press 2011)

Although all of these platforms enable users to create complex, spatially dynamic models, each has strengths and weaknesses regarding user-friendliness and ability to handle large models. One advantage of model development in Netlogo is the platform's built-in 'system dynamics modeler,' which translates SD models into Netlogo code. Additionally, *REPAST Simphony*, an interactive, cross-platform modeling environment can also now import Netlogo models, allowing users to rapidly develop models in Netlogo (with minimal technical expertise[4]), and execute them in *REPAST's* high performance computing environment (often necessary for large, spatial simulations).

[Insert Figure 3**]**

[4] *Netlogo* models are relatively easy to develop compared to the JAVA programming required for traditional REPAST models.

We begin with a simple model whereby node dynamics are uniformly defined as a single stock (S_i) that slowly grows exponentially due to influence from neighboring nodes ($S_{N(i)}$):

$$S_{i_{t+1}} = S_{i_t} + .001 * \sum_{j \in N(i)} S_{j_t}$$

Panel A of Figure 3 shows the initial network graph, randomly generated by Netlogo. The model was then run for an arbitrary period of steps, resulting in new stock values for each node, yielding Panel B of Figure 3. The nodes that are highly connected consequently receive disproportionate share of the system wide growth compared to the nodes of low degree, because of the spatial feedback. Thus, an immediate issue is the visualization associated with the stock within each node, which we decided to depict as:

$$NodeSize_i = 0.1 + \sqrt{\frac{S_i}{Avg(S_{j \in N(i)})}}$$

This visualization could be modified to depict actual stock sizes, although this can quickly preclude continued visualization within the same network topology (i.e. each node grows to large to show).

In our second example, we implement a simple disease spread model, commonly known as an SIR (susceptible-infected-recovered populations) model (Homer and Hirsch 2006). These models are common in the epidemiological literature (Capasso 1993) and have been translated into the SD framework in various instances (Ritchie-Dunham 1999; Sterman 2000; Rich 2008).

Like the previous example, we begin with a random graph representing connections between different nodes (e.g. road connections between neighboring towns; Figure 4a). Within each node, an individual SIR model operates (Figure 4b), diffusing sick individuals into nearby nodes based on diffusion rate *d*, the number of sick individuals in the surrounding nodes (I_v; v is the neighborhood set of *i.*), and the number of susceptible individuals in the target node (S_i). As shown in the equation below, the diffusion rate (*d*) modifies the infection rate (r_f). The number of sick individuals in the target node is also influenced by the infection rate (r_f) multiplied by the susceptible (S_i) and infected (I_i) proportions of the population (P_i) and the rate of recovery (r_r).

$$I_{i_{t+1}} = I_{i_t} + \left(d \frac{\sum_{j\in N(i)} I_{j_t}}{\sum_{j\in N(i)} P_{j_t}} + r_f \frac{I_{i_t}}{P_{i_t}} \right) S_{i_t} - r_r I_{i_t}$$

The infection (signified by squares) begins near the lower right corner (Panel C), spreading faster to more highly connected nodes (Panel D), eventually hitting the upper left corner (further away, as measured by network distance), but completely missing non-connected nodes (see pocket of nodes in lower left, and two individual nodes on right side of graph). After the infection has swept through the network (Panel D), infected individuals begin to recover (triangles), which sweep through the network as another wave (Panel E). An aggregate measure of the infected and recovered populations mimics classic SIR model behavior (Panel F; Sterman 2000).

Finally, we demonstrate a more complex example involving a dynamic network (Figure 5). In many cases, dynamic networks can add nearly infinite complexity to models (see Breiger et al. (2003) and Metcalf and Paich (2005) for an exploration of the spatial-dynamics of social networks). The network representations can be easily made dynamic, simply by adding binary weights allowing us to represent links as binary connections (e.g. on/off, social connection/no social connection) that can change over time, or even as a continuum of values of non-zero weights (e.g. acquaintances, friends, good friends, spouse, etc.), which may define the strength and frequency of interaction), which is important for representations such as SIR models. These weights can change over time either independently or conditioned on the attributes of the nodes the links connect.

For example, instituting a quarantine policy (e.g. triggered when the infected population within a node reaches > 30%) that attempts to shut down disease diffusion by eliminating links will drastically alter the spread and recovery pattern (e.g. Figure 5d. In our example, the links are restored when the infected population proportion is less than 10% (see Figure 5e). Therefore, the space itself co-evolves with the underlying dynamic processes, thus better representing the complex dynamics of quarantine policies and their spatial effects.

Conclusions and Discussion

Spatial system dynamics models are not new. However, close attention has not been paid to the representation of space in these models. Contrasting the rigorous, scientific process of defining causal mechanisms in dynamic systems, little thought seems to be given to how and why we represent space in SD models.

This article pursues a unified, theoretical underpinning to inform how and why we represent space in system dynamics models. To do this, we portray spatial processes in two different ways. First, we can characterize a spatial process as an extensive process if we are purely concerned with its behavior at the boundary of a given space (e.g. if a product has entered a market, if a disease has entered a village, if a city has reached a certain density in a given neighborhood). Conversely we can characterize intensive processes as processes where we are concerned with how spatial structure affects the process dynamics within the boundary.

Recent spatial-system dynamics research has articulated ‘space’ as a tessellation into regular grids (Ahmad and Simonovic 2004; BenDor and Metcalf 2006). The increasingly common, yet simple, tessellation of underlying space into grids whereby individual processes affecting one another is only one way to representation space. Similar tessellation can employ hexagons, triangles, and other geometric shapes. However, recent research has shown that this process is highly susceptible to artifacts of grid geometry (Chen and Pontius In Press), which is likely to go undetected in SD modeling. It is extremely difficult to perform sensitivity analyses on grid resolution and size, particularly when spatial data is available at low spatial resolution.

We argue that in order to abstract away the artifacts of this tessellation, we should instead view spatial interactions as they occur across a topological network that defines the underlying structure of space. By articulating space through networks, we can abstract away arbitrary grid representations and more rigorously (and easily) study how models are affected by particular spatial representations.

The weighted network model that we discuss in our final example endows attributes to both nodes and links, allowing us to model the co-evolution of space alongside dynamic disease processes. This contrasts with raster based SD models, where spatial pattern is determined by collecting the homogenous values of the processes within a grid, requiring that the underlying spatial structure remains invariant. The network representation of space treats the spatial relationships themselves as dynamic and therefore allows for changes in the local spatial structure affecting the global process dynamics.

Networks also facilitate construction and use of irregular tessellations of space, accommodating diverse spatial representations, including raster and vector models of landscapes, social connections and networks, and diffusion vectors. Under a network representation, grids can be represented as a regular graph (on a torus[5]) or a near regular graph (on a plane; such as Figure 2b). Polygons are intuitive for depicting heterogeneous spaces, and are the standard representation of political boundaries such as cities, counties and countries. Similarly, lines are intuitive representation of geographical phenomena such as rivers, or infrastructure such as roads and water networks. Non-contiguous regions can have spillover effects on their distance-based

[5] This could be used if edge effects are particularly problematic.

neighbors. Even in contiguous regions, spillover effects may be due to second-order neighborhood relationships or relationships that vary in strength (weighted relationships). All these common issues and concerns over spatial characterization can be unified under standard network topology.

Building on years of visualization research in aspatial SD (and other fields, including computer graphics; Dykes 1997), future research should also explore spatial-dynamic visualization techniques. Extensive processes result in archetypical spatial patterns such as linear growth and oscillations. In Table 2, we depict examples of very simple archetypical spatial behavior and potential modes of visualization. However, intensive processes are not, in our experience, easily amenable to such visual representations. Extending models spatially means abandoning common, 2-D graphical visualizations of the behavior of system elements. Rather, methods and software need to be developed for exploring 4-D or 5-D (3-dimensions, time, and value) representation of maps and networks.

As the system dynamics method evolves and becomes more sophisticated, strong theories informing model spatialization and the spatial-dynamic modeling process will become increasingly important. Many of the considerations that currently introduce rigor into the SD modeling process, including the use of historical behavior as reference mode information, dynamic hypothesis creation, and iteration in the model construction process, have spatial analogues. The same rigor should be used in 1) determining spatial representations (zonal, gridded, vector, network, etc.), 2) thinking through archetypical spatial processes (e.g. density dependent growth and resulting diffusion; BenDor and Metcalf 2006). Expanding the scientific basis of SD into the spatial realm will enrich both the SD and spatial science and enable modelers to create more accurate, useful, and usable spatial-dynamic models.

References

Ahmad, S. and Simonovic, S.P., 2004. Spatial System Dynamics: New Approach for Simulation of Water Resources Systems. *Journal of Computing in Civil Engineering*, 18(4):331-340.

Alekseeva, E.I. and Kirzhner, V.M., 1994. Migration on networks and its stability consequences. *System Dynamics Review*, 10(1):63-85.

Allen, T. and Hoekstra, T.W., 1993. *Toward a unified ecology*. New York, NY: Columbia University Press

Anselin, L., 2002. Under the Hood: Issues in the Specification and Interpretation of Spatial Regression models. *Agricultural Economics*, 27(3):247-267.

Anselin, L., 2003. *GeoDa™ 0.9 User's Guide* [Online]. Urbana, IL: Center for Spatially Integrated Social Science. Available from: http://geodacenter.org/downloads/pdfs/geoda093.pdf [Accessed: 1/2/2011].

Banks, J., Carson, J., Nelson, B.L. and Nicol, D., 2004. *Discrete-Event System Simulation (4th Edition)*. New York, NY: Prentice Hall

BenDor, T. and Metcalf, S., 2006. The Spatial Dynamics of Invasive Species Spread. *System Dyanmics Review*, 22(1):27-50.

Bhat, U.N. and Miller, G.K., 2002. *Elements ofApplied Stochastic Processes (Third Edition)*. New York, NY: Wiley and Sons

Borshchev, A. and Filippov, A., 2004. *From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools* [Online]. Oxford, England: System Dynamics Society. Available from: http://www.systemdynamics.org/conferences/2004/SDS_2004/PAPERS/381BORSH.pdf [Accessed: 1/2/2011].

Breierova, L., 1997 *Generic Structures: Overshoot and Collapse (DD-4480)* [Online]. Cambridge, MA: MIT System Dynamics Education Project Sloan School of Management. Available from: http://sysdyn.clexchange.org/sdep/Roadmaps/RM9/D-4480.pdf [Accessed: 1/2/11].

Breiger, R., Carley, K. and Pattison, P. (Editors), 2003. Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers. National Research Council Committee on Human Factors, Washington, D.C.

Burdekin, R., 1979. A dynamic spatial urban model: A generalization of Forrester's urban dynamics model. *Urban Systems*, 4(2):93-120.

Capasso, V., 1993. *Mathematical structures of epidemic systems (2nd Edition)*. New York, NY: Springer

Chen, H. and Pontius, R.G., In Press. Sensitivity of a Land Change Model to Pixel Resolution and Precision of the Independent Variable. *Environmental Modeling and Assessment*.

Chung, C.V., 2001. *Generic Structures in Oscillating Systems I (D-4426-3)* [Online]. Cambridge, MA: MIT System Dynamics Education Project Sloan School of Management. Available from: http://sysdyn.clexchange.org/sdep/Roadmaps/RM6/D-4426-3.pdf [Accessed: 1/2/2011].

Collier, N.T. and North, M.J., In Press 2011. *Repast SC++: A Platform for Large-scale Agent-based Modeling*. *In*: K.K. W. Dubitzky, and B. Schott (Editor), *Large-Scale Computing Techniques for Complex System Simulations*. New York, NY: Wiley and Sons.

Couclelis, H., 1992. *People manipulate objects (but cultivate fields): Beyond the raster-vector debate in GIS*. *In*: A.U. Frank, I. Campari and U. Formentini (Editor), *Theories and* *methods of spatio-temporal reasoning in geographic space*. Berlin, Germany: Springer Verlag., 65-77.

Cox, D.R. and Isham, V., 1980. *Point processes*. Boca Raton, FL: CRC Press

Cruz, J.P. and Olaya, C., 2008. *A System Dynamics Model for Studying the Structure of Network Marketing Organizations* [Online]. Athens, Greece: System Dynamics Society. Available from: http://www.systemdynamics.org/conferences/2008/proceed/papers/OLAYA411.pdf [Accessed:

Diestel, R., 2006. *Graph Theory*. New York, NY: Springer-Verlag

Dykes, J.A., 1997. Exploring spatial data representation with dynamic graphics. *Computers and Geosciences*, 23(4):345-370.

Egenhofer, M.J., Glasgow, J., Gunther, O., Herring, J.R. and Peuquet, D.J., 1999. Progress in computational methods for representing geographical concepts. *International Journal of Geographical Information Systems*, 13(8):775-796.

Ford, A., 2009. *Modeling the Environment (2nd Edition)*. Washington, D.C.: Island Press

Forrester, J., 1968. Market Growth as Influenced by Capital Investment. *Sloan Management Review*, 9(2):83-105.

Forrester, J.W., 1969. *Urban Dynamics*. Waltham, MA: Pegasus Communications

Galton, A., 2001. Space, Time, and the Representation of Geographical Reality. *Topoi* 20(2):173-187.

Glick, M. and Duhon, T., 2001. *Generic Structures: S-Shaped Growth I* [Online]. Cambridge, MA: MIT System Dynamics Education Project Sloan School of Management. Available from: http://sysdyn.clexchange.org/sdep/Roadmaps/RM5/D-4432-2.pdf [Accessed: 1/2/2011].

Goodchild, M., 1992. Geographic Data Modeling *Computers and Geosciences*, 18(4):401-408.

Homer, J.B. and Hirsch, G.B., 2006. System Dynamics Modeling for Public Health: Background and Opportunities. *American Journal of Public Health*, 96(3):452-458.

Lane, D.C., 1998. Can we have confidence in generic structures? *Journal of theOperational Research Society*, 49(936-947.

Lee, D.B., 1973. Requiem for Large-Scale Models. *Journal of the American Planning Association*, 39(3):163-178.

Longley, P., Goodchild, M.F., Maguire, D.J. and Rhind, D.W., 2005. *Geographic information systems and science (2nd edition)*. New York, NY: John Wiley and Sons

Lowry, T. and Taylor, R., 2009. *Visualizing Spatial and Temporal Dynamics in Google Earth Using Powersim* [Online]. Albuquerque, New Mexico: Available from: http://www.systemdynamics.org/conferences/2009/proceed/papers/P1405.pdf [Accessed: 1/2/2011].

Maxwell, T. and Costanza, R., 1997a. A Language for Modular Spatio-Temporal Simulation. *Ecological Modelling*, 103-113(105.

Maxwell, T. and Costanza, R., 1997b. An Open Geographic Modeling Environment. *Simulation*, 68(3):175-185.

McGarigal, K. and Marks, B., 1995. *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure (PNW-GTR-351)* [Online]. Portland, OR: Pacific Northwest Research Station, US Department of Agriculture. Available from: http://www.fs.fed.us/pnw/pubs/pnw_gtr351.pdf [Accessed: 1/2/2011].

Metcalf, S. and Paich, M., 2005. *Spatial Dynamics of Social Network Evolution* [Online]. Boston, MA: System Dynamics Society. Available from: http://www.systemdynamics.org/conferences/2005/proceed/papers/METCA487.pdf [Accessed: 1/2/2011].

Mosekilde, E., Aracil, J. and Allen, P.M., 1988. Instabilities and chaos in nonlinear dynamic systems *System Dynamics Review*, 4(1-2):14-55.

Paich, M., 1985. Generic Structures. *System Dynamics Review*, 1(126-132.

Pfaffenbichler, P., Emberger, G. and Shepherd., S., 2010. A system dynamics approach to land use transport interaction modelling: the strategic model MARS and its application. *ystem Dynamics Review*, 26(3):262-282.

Pourdehnad, J., Maani, K. and Sedehi, H., 2002. *System Dynamics and Intelligent Agent-Based Simulation: Where is the Synergy?* [Online]. Palermo, Italy: System Dynamics Society. Available from: http://www.systemdynamics.org/conferences/2002/proceed/papers/Pourdeh1.pdf [Accessed: 1/2/25011].

Reggiani, A. and Nijkamp, P., 1995. Competition and complexity in spatially connected networks. *System Dynamics Review*, 11(1):51-66.

Rich, K.M., 2008. An interregional system dynamics model of animal disease control: Applications to foot-and-mouth disease in the Southern Cone of South America. *System Dynamics Review*, 24(67-96.

Richardson, G.P., 1999. *Feedback Thought in Social Science and Systems Theory*. Waltham, MA: Pegasus Communications

Ritchie-Dunham, J.L., 1999. Evaluating epidemic intervention policies with systems thinking: A case study of dengue fever in Mexico. *System Dynamics Review*, 15(2):119-138.

Saeed, K., 1992. Slicing a Complex Problem for System Dynamics Modeling. *System Dynamics Review*, 8(3).

Saeed, K., 1998a. Defining a Problem or Constructing a Reference Mode. The 16th International Conference of The System Dynamics Society, Quebec City, Canada, July 20 - 23, 1998, http://www.systemdynamics.org/conferences/1998/PROCEED/00016.PDF.

Saeed, K., 1998b. Defining Developmental Problems for System Dynamics Modeling: An Experiential Learning Approach. Worcester Polytechnic Institute Department of Social Science and Policy Studies Working Paper No. 26.

Saeed, K.S.D.S., 2001. Defining developmental problems for policy intervention or Building reference mode in 20 steps over 5 learning cycles. Proceedings of the 19th International Conference of System Dynamics Society. July 23-27, 2001. Atlanta, GA.

Sterman, J., 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. New York, NY: Irwin/McGraw-Hill

Ulli-Beer, S., Gassmann, F., Bosshardt, M. and Wokaun, A., 2010. Generic Structure to Simlate Acceptance Dynamics. *System Dynamics Review*, 26(2):89-116.

Voinov, A., Costanza, R., Fitz, C. and Maxwell, T., 2007. Patuxent Landscape Model: 1. Hydrological Model Development. *Water Resources Research*, 34(2):163-170.

Voinov, A., Costanza, R., Wainger, L., Boumans, R., Villa, F., Maxwell, T. and Voinov, H., 1999. The Patuxent landscape model: Integrated ecological economic modeling of a watershed. *Environmental Modeling and Software*, 14(473-491.

Wilensky, U., 1999. *NetLogo* [Online]. Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University. Available from: http://ccl.northwestern.edu/netlogo/. [Accessed:

Wils, W., 1974. *Metropolitan Population Growth, Land Area, and the Urban Dynamics Model*. *In*: N.J. Mass (Editor), *Readings in Urban Dynamics*. Cambridge, MA: Wright-Allen Press, 103-114.

Wolfram, S., 1983. Statistical Mechanics of Cellular Automata. *Reviews of Modern Physics*, 55(601-644.

Wolstenholme, E.F., 2003. Towards the definition and use of a core set of archetypal structures in system dynamics. *System Dynamics Review*, 19(1):7-26.

Wolstenholme, E.F., 2004. Using Generic System Dynamics Archetypes to Support Thinking and Modelling. *System Dynamics Review*, 20(4):341-356.

Table 1: Examples of Non-Spatial Systemic Archetypes (k=constant or adjustment time [k_i, k_o, k_R, k_S = inflow, outflow, R-related, and S-related constants, and, S, R=stocks, C=goal, carrying capacity, or 'normal condition' [C_r=R-related goal or normal condition])

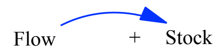
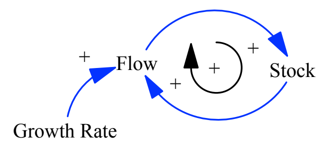
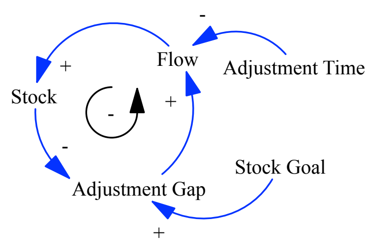
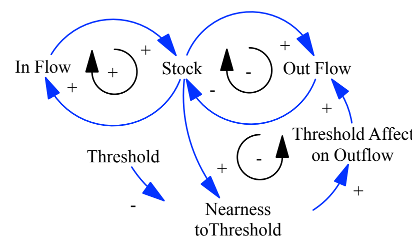
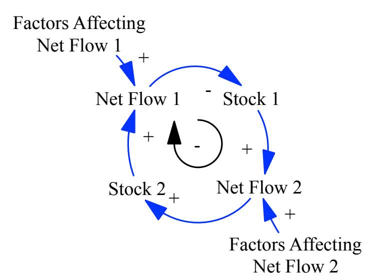
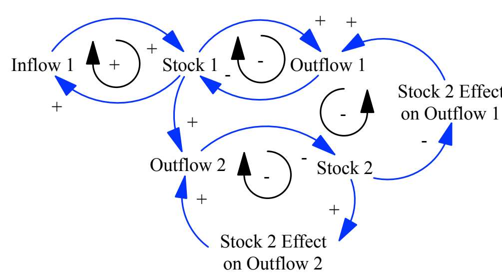
Systemic Archetype	**Governing Equations**	**Causal Loop Diagram**
1) Linear growth	$\frac{dS}{dt} = k$	
2) Exponential growth	$\frac{dS}{dt} = kS$	
3) Goal seeking growth	$\frac{dS}{dt} = \frac{C-S}{k}$	
4) Logistic growth	$\frac{dS}{dt} = kS(1-\frac{S}{C})$	
5) Sustained oscillations	$\frac{dS}{dt} = k_S R$ $\frac{dR}{dt} = k_R S$	
6) Overshoot and collapse	$\frac{dS}{dt} = k_i S - k_o S\frac{R}{C_R}$ $\frac{dR}{dt} = -k_R S$	

Table 2: Examples of spatial systemic archetypes where patterns are specified for change in the area of a circle (*not the radius*; k=constant or adjustment time [k_i, k_o, k_R, k_S = inflow, outflow, R-related, and S-related constants, and, S, R=stocks, C=goal, carrying capacity, or 'normal condition' [C_r=R-related goal or normal condition])

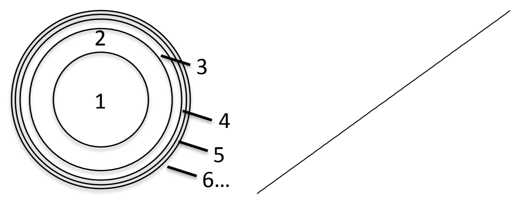
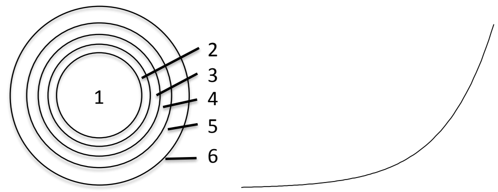
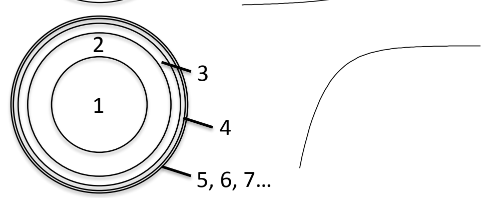
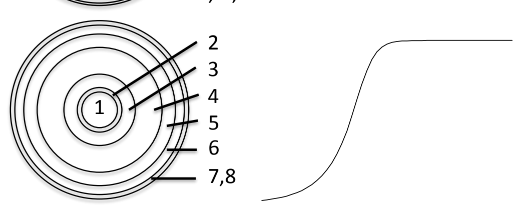
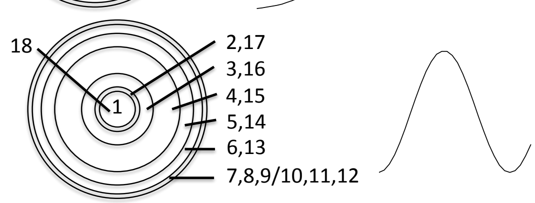
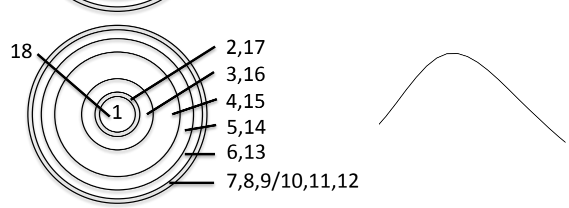
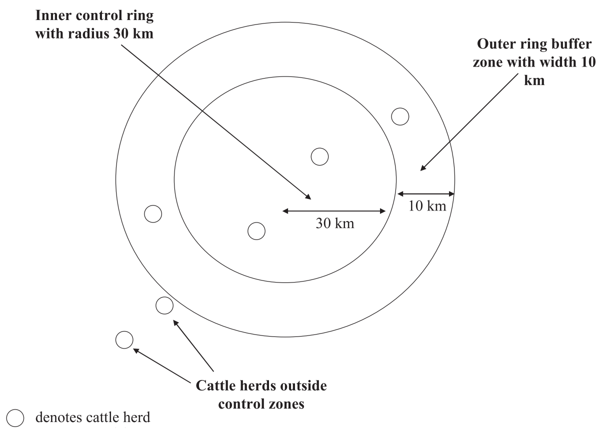
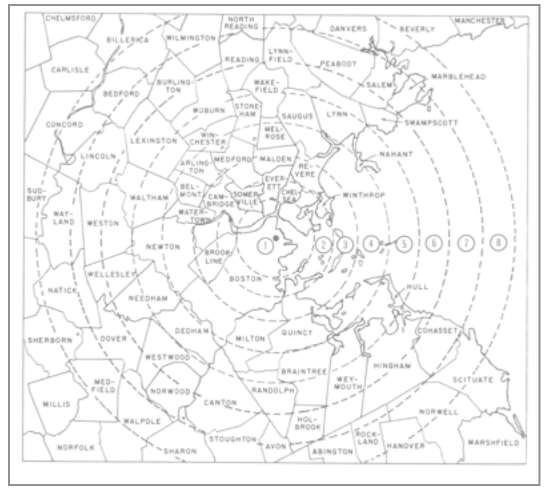
Systemic Archetype	**Governing Equations**	**Spatial Dynamic Visualization**
1) Linear growth	$\frac{dr}{dt} = \frac{k}{2\pi r}$	
2) Exponential growth	$\frac{dr}{dt} = \frac{kr}{2}$	
3) Goal seeking growth	$\frac{dr}{dt} = \frac{C}{2\pi kr} - \frac{r}{2k}$	
4) Logistic growth	$\frac{dr}{dt} = \frac{kr}{2}(1 - \frac{\pi r^2}{C})$	
5) Sustained oscillations	$\frac{dr_1}{dt} = \frac{kr_2^2}{2r_1}$ $\frac{dr_2}{dt} = \frac{kr_1^2}{2r_2}$	
6) Overshoot and collapse	$\frac{dr_1}{dt} = r_1(k_i - \frac{k_o \pi r_2^2}{C_R})$ $\frac{dr_2}{dt} = -\frac{k_R r_1^2}{2r_2}$	

Figure 1: Examples of Spatial Representation in System Dynamics Models.

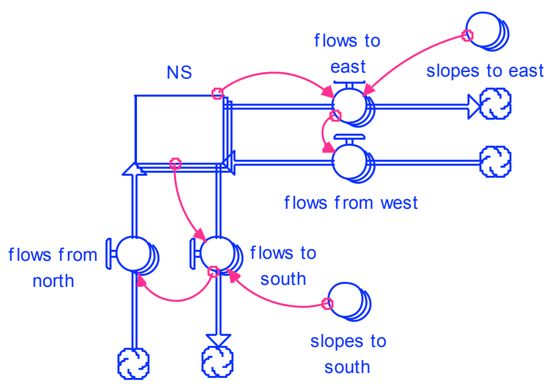
Panel A: Local spatial spread zones in Rich (2008) model of South American Foot-and-Mouth Disease spread.



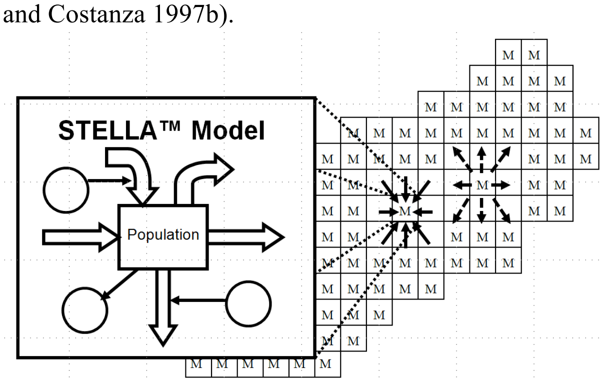
Panel B: Wils (1974) zonal extension of the Forrester (1969) Urban Dynamics model.



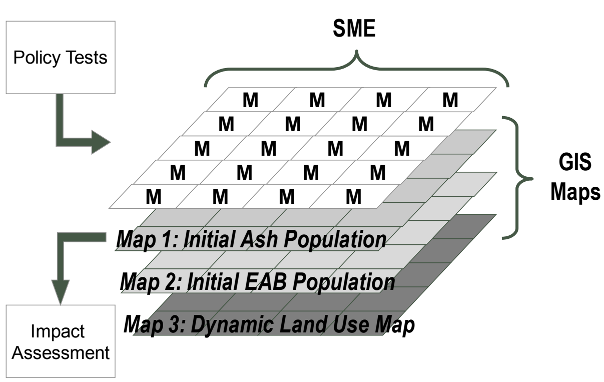
Panel C: Ford (2009) model of water flowing through a drainage basin.



Panel D: Spatial Modeling Environment (SME) implementation of SD models in each grid cell (Maxwell and Costanza 1997b).



Panel E: BenDor and Metcalf (2006) invasive species spread (Emerald Ash Borer) model, implemented in SME.



Panel F: Hydrologic routing schemes used to model water moving (a) from one cell to the next one, (b) over several cells in one time step, and (c) under variable path length algorithm, the amount of water in the donor cell determines how far it travels. From Voinov et al. (2007) Patuxent watershed landscape model.

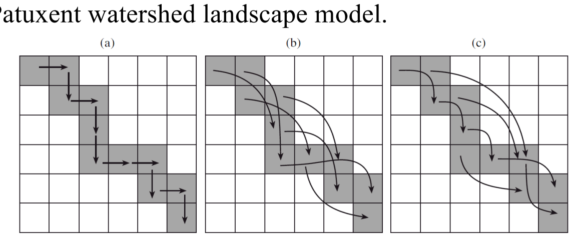
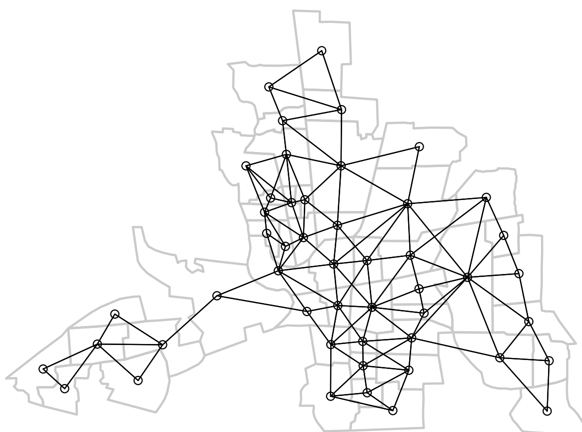
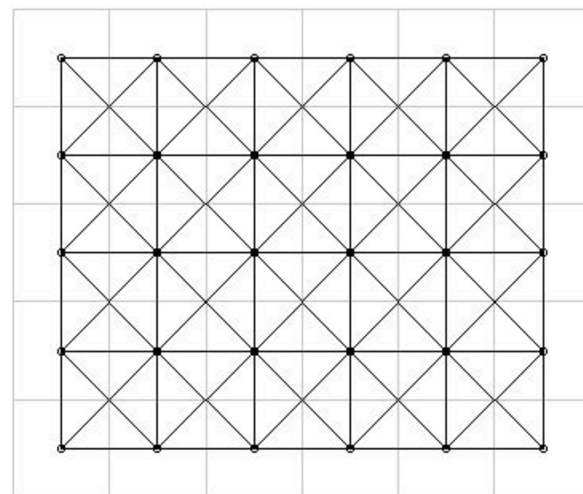


Figure 2: Network Representations of Space

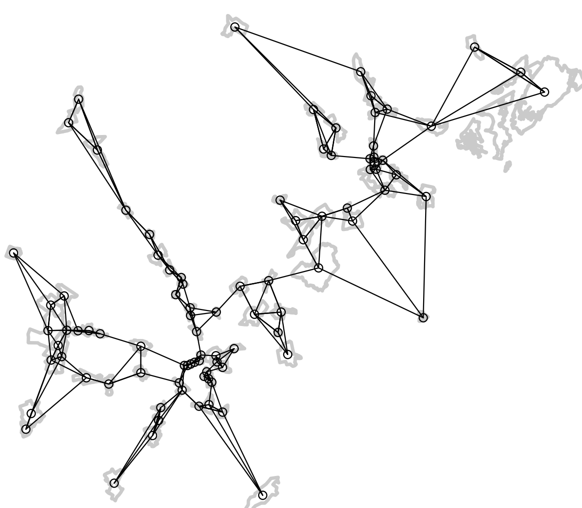
Panel A: Network representation of complex, non-uniform polygon map (Columbus, Ohio neighborhoods; Anselin 2003)



Panel B: Nearly 'regular' graph as network representation of a grid – each node is equally connected to all contiguous neighbors



Panel C: Non-contiguous neighborhood connections among spatially disconnected parcels.



Panel D: Example of 'second order' connections, where polygons are connected to all neighbors of their neighbors. Note that the number of connections have increased geometrically from that of Panel A.

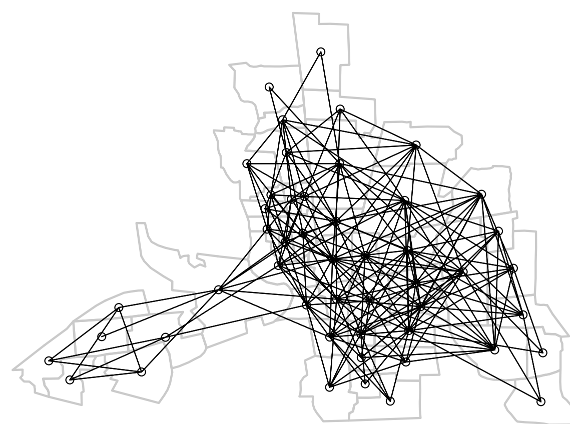


Figure 3: Intensive Process Examples on Random Networks. **Panels A-B:** Size indicates each node's relative stock size as determined from the size of the stocks in connected nodes.

Panel A: Exponential Growth Initialization Panel B: Exponential Growth Results

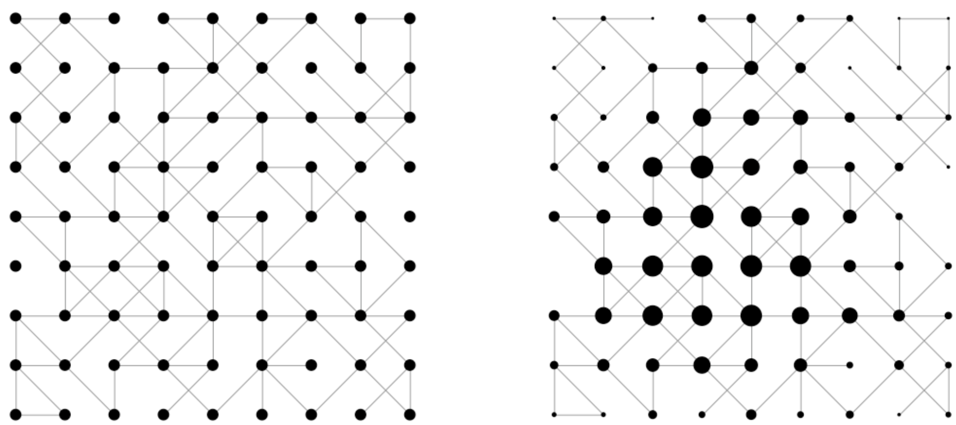
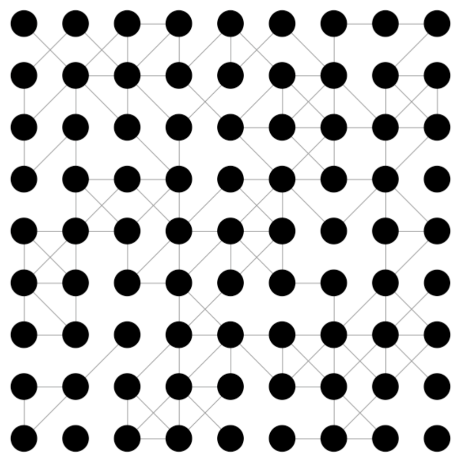
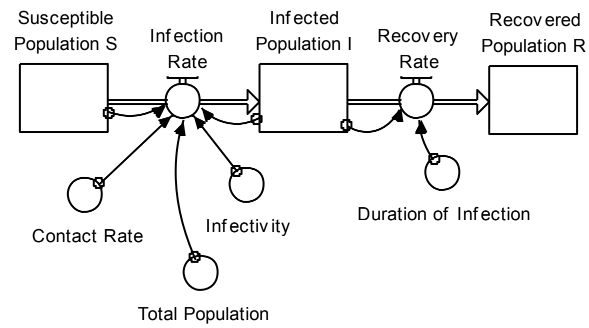


Figure 4: Network Representation of SIR Model. Shapes determine dominant type of population; circles indicate susceptible (Panel A), squares indicate infected, and triangles indicate recovered populations that dominate the node.

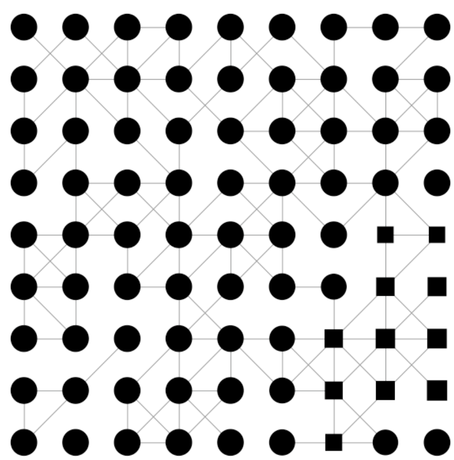
Panel A: SIR Model Initialization



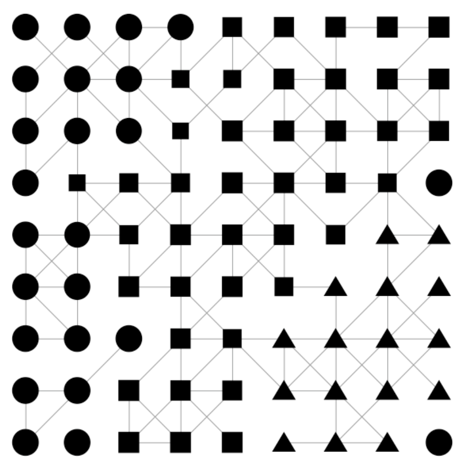
Panel B: Classic SIR SD Model (Sterman 2000)



Panel C: SIR Model (Timestep = 10)

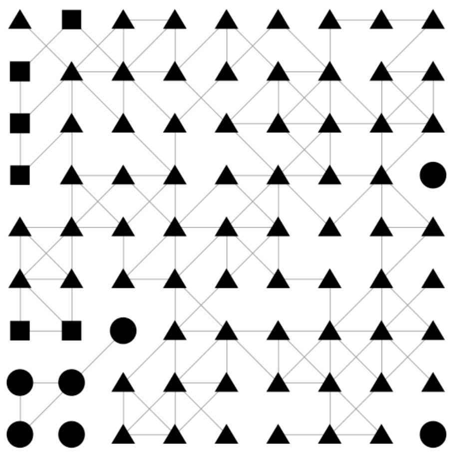


Panel D: SIR Model (Timestep = 20)



Panel E: SIR Model (Timestep = 30)

Panel F: Aggregate dynamic pattern of total population (all nodes)



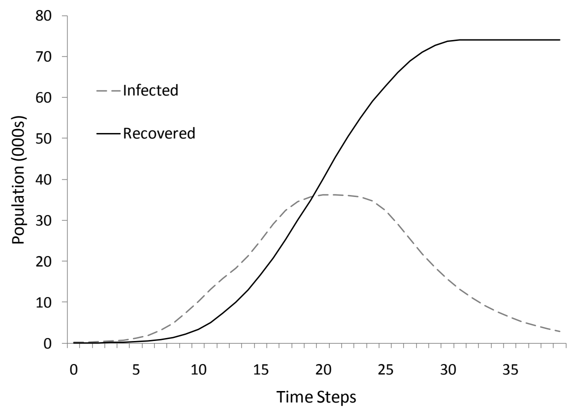
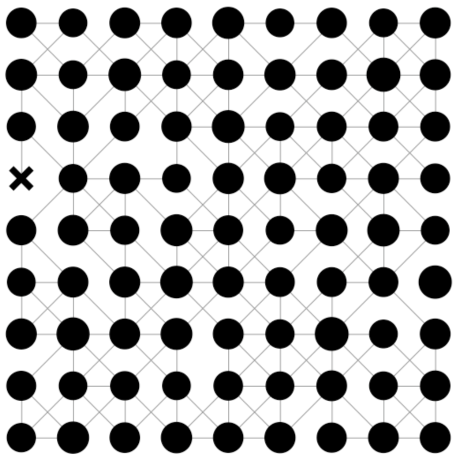
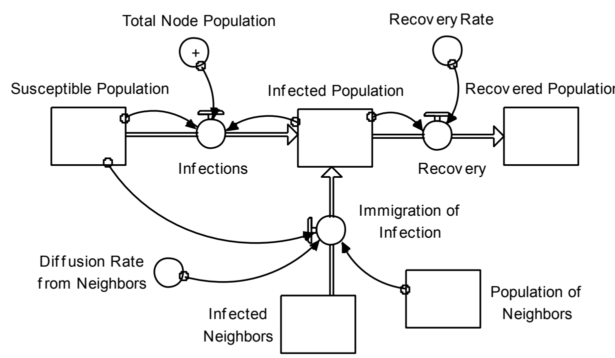

Infected
Recovered
Population (000s)
Time Steps
0
10
20
30
40
50
60
70
80
0
5
10
15
20
25
30
35

Figure 5: Dynamic Network Representation of SIR Model. Shapes determine dominant type of population (size determines relative number); circles indicate susceptible, squares indicate infected (infection origin noted with 'X'), and triangles indicate recovered populations that dominate the node.

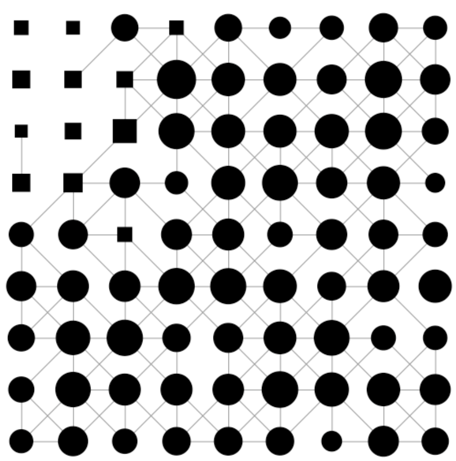
Panel A: SIR Model Initialization



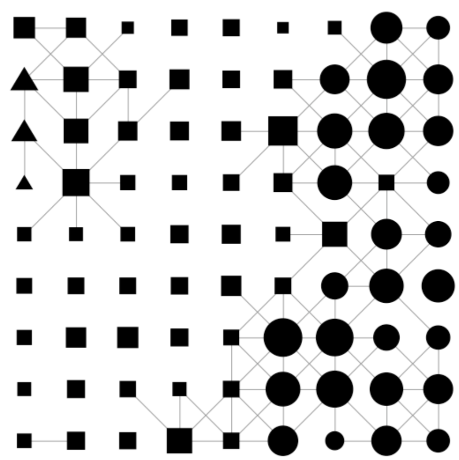
Panel B: Classic SIR SD Model (Sterman 2000)



Panel C: SIR Model (Timestep = 20)

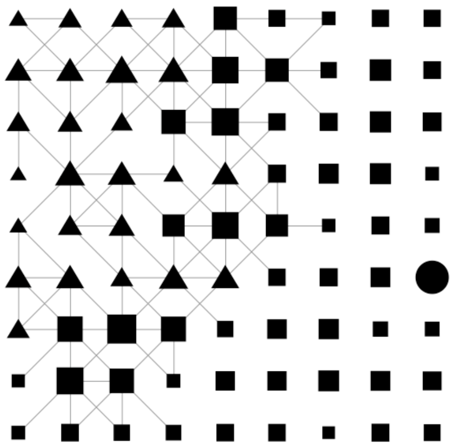


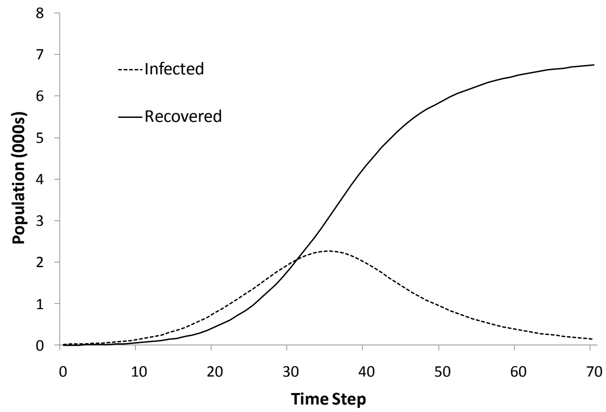
Panel D: SIR Model (Timestep = 30)



Panel E: SIR Model (Timestep = 40)

Panel F: Aggregate dynamic pattern of total population (all nodes)




Infected
Recovered
Population (000s)
Time Step
0
1
2
3
4
5
6
7
8
0
10
20
30
40
50
60
70