Full Spectrum Modeling: From Simplicity to Elaboration and Realism in Urban Pattern Formation

William Rand
Northwestern University
wrand@northwestern.edu

Uri Wilensky
Northwestern University
uri@northwestern.edu

Abstract
There is a dichotomy between researchers who utilize elaborated and realistic agent-based models and those who prefer simple models. Researchers who use realistic agent-based models (ABMs) argue that their realism makes them more useful, while those who utilize simple models claim that they are more general and provide more global knowledge. We believe that it is better to embrace both of these modeling approaches at the same time. Creating richly, detailed, empirically validated models to explore complex systems at a high level, and at the same time extracting general principles to create simplified models that help explore the most important general principles of the phenomena. We call this approach full spectrum modeling and we examine its use through a particular project concerned with understanding urban pattern formation.

Contact:
Dr. William Rand
Northwestern Institute on Complex Systems (NICO)
Northwestern University
600 Foster St.
Evanston, IL 60208

Tel: 1-847-491-5734
Fax: 1-847-467-1280
Email: wrand@northwestern.edu

Key Words: agent-based modeling, city development, pattern formation, urban modeling, simple models, realistic models

Support: This paper is part of the Cities project and was supported in part by the National Science Foundation (NSF) ITR award 0326542 under the direction of Dr. Uri Wilensky. The Northwestern Institute on Complex Systems (NICO) provided support for William Rand.
Full Spectrum Modeling: From Simplicity to Realism in Urban Pattern Formation
William Rand and Uri Wilensky

Introduction

Many agent-based models (ABMs), like Schelling’s original segregation model [Schelling, 1978], are very simple, and critics often claim that they do not represent real world environments. On the other hand, these simple models often illustrate very general principles about the world that can best be understood through a “caricature” of the real world, which is exactly what these models provide. For instance, in the case of Schelling’s model, the simplified version of the world illustrates quite well that it is not required that people be explicitly racist to create segregated neighborhoods. Thus though simple models lack every mechanism that exists in the real world they do provide us with tools for understanding complex phenomena.

Elaborated, realistic models, on the other hand, that contain a variety of different mechanisms and use empirical data, can in fact be shown to match the actual world. One could imagine a more elaborated version of Schelling’s model that read in real data from Chicago neighborhoods examining what people’s preferences were, and used cognitive modeling to determine how they would move around the neighborhood. Within the context of a complex system, realistic models can be used to explore particular situations and particular problems. However, some critics charge these models are impossible to understand, and that they apply only to the particular details that they are built around [Grimm et all, 2005]. But sometimes that is exactly the desired outcome, a particular prediction. For instance, a realistic Schelling model might be able to predict land development patterns in neighborhoods of Chicago for years to come.

This conflict between elaborated, realistic models and simple models has been discussed both explicitly and implicitly by researchers [Axelrod, 1997; Carley, 2002]. For instance, Axelrod [1997] argues that simple models provide for better transparency and a deeper level of understanding. On the other hand, Carley [2002] argues for an increased emphasis on elaborated and realistic models since they are more consistent with the scientific method and more easily falsifiable. However, this forced choice between simple models and realistic models is not required. Grimm et al [2005], through their pattern-oriented modeling (POM) approach, argue that instead the optimal level of modeling complexity resides somewhere in the middle. This approach allows a single model to reproduce multiple levels of observed patterns. While we agree that being able to create patterns at multiple levels of hierarchy is important, we argue that by restricting the process to building one model at the “sweet spot” of complexity the POM approach falls short. Instead, in this paper, we argue for building models at different levels of complexity, and thus exploring both general principles and specific applications at the same time. We call this approach full spectrum modeling, and, as the name implies, it involves modeling phenomena at multiple levels of detail. Not just modeling at the level of “simple” and “elaborated and real” but also at the levels in-between. We argue that a full spectrum approach allows for a deeper and fuller understanding of the given phenomenon than any single-level approach.

We will begin this paper by exploring the costs and benefits of simple and elaborated-realistic models (ER models). We will then introduce the approach we call full spectrum modeling and move beyond just the combination of these two single-level approaches to show how full spectrum modeling provides synergistic benefits. We will examine this modeling approach specifically within the application area of urban formation.

What is Simple? What is Elaborated and Realistic?

Before we begin describing the benefits of different modeling approaches and in particular the full spectrum approach, it is important to clarify what we mean by simple and elaborated-realistic models. As mentioned, simple models allow for the examination of general principles in detail. When we talk about simple models we are referring to models that our simple from two different perspectives. First, these models can be simple because they have no necessary connection to real world / empirical data. Instead they are based on hypothetical characteristics and theoretical constructs about the world. Second, these models are simple because they have few types of agents

---

1 Sometimes this can take the form of a series starting with simple models and becoming progressively realistic, or starting with realistic models and becoming progressively simplified.
and few mechanisms operating. Instead they may assume that all agents are of a similar type or they exaggerate one particular mechanism and examine it in detail\(^2\).

Thus ER models are more complicated from these two perspectives. First they are often based on realistic data and results. They incorporate data derived from GIS systems, or social network analysis, for instance. Second, ER models utilize many different types and levels of agents, each of which has a wide variety of actions that they can carry out. These different mechanisms often seem to make the model more realistic and allow the agents to carry out many more options.

We will be talking about simple versus ER models, but as full spectrum modeling implies we really mean a whole spectrum of models from the simplest to the most elaborated and realistic. However, discussing the two extremes of this spectrum will help us illustrate the benefits of such an approach.

When we discuss simple models versus ER models, it is important to note that we are referring to models that are simple from the design standpoint. A good test of this is how long it takes to describe the model at a procedural level. If it takes many pages of description and lots of examples then the model is probably not simple. However, just because the design of these models is simple does not mean that the results of these models are easy to understand. The results of these models could still be quite elaborate. For instance, the Ising model\(^3\) though very simple to describe has a lot of complex behavior and thus the results of this model are quite elaborate [McCoy and Wu, 1973]. It is also possible to have a ER model that has simple results. If an ER model returns the same results no matter what the inputs and how the mechanisms are manipulated then its output is very simple. An apparent example is the Club of Rome model which regardless of the inputs to the model results in the conclusion that the world will run out of resources [Meadows, Meadows, Randers, and Behrens, 1972].

**The Benefits and Costs of Simple Models**

Proponents of simple models claim that such models illustrate demonstrative general principles about the world. Thus they go beyond any particular set of real-world data, and ignore specific mechanisms precisely so that they can be understood at a more general level. In fact if the model was easily calibrated to a particular real-world instance then it would lose its generalizability and thus would only apply to the particular situations being addressed. Thus simple models capture the essence of the systems being modeled and provide a more clear explanation than ER models do. For instance, in Axelrod and Hammond’s well-known ethnocentrism model [2003], the general principle is that, regardless of a large variety of parameter settings, ethnocentric behavior will evolve and persist.

By ignoring many mechanisms that exist in a system, simple models provide a “caricature” of the system that exaggerate one particular aspect so that it can be examined in-depth. All models must simplify to some extent or another since they cannot hope to capture all of the interactions going on in the real world. Thus there will always be mechanisms that are left out, and simple models choose to make only one or two of those aspects prominent. This enables us to examine a particular mechanism at a high level of resolution. For instance, the ethnocentrism model concentrates mainly on the evolutionary process and how local reproduction events give rise to ethnocentric behavior. This focus allows Axelrod and Hammond to examine exactly what properties of this mechanism are critical to the perpetuation of ethnocentrism. Thus simple models can actually give us a more detailed understanding of particular mechanisms than ER models can, since we can explore a few mechanisms in isolated circumstances.

On the other hand, models that contain few mechanisms and lack real-world data are often dismissed as being simple and unrealistic. Stakeholders who have a particular interest in a modeling project will claim that simple models ignore important real world details. For instance, the ethnocentrism model lacks any notion of institutional restraint. Sometimes the complaint of missing mechanisms is made purely because the stakeholder has a personal interest in making sure their favorite topic is treated appropriately, but sometimes this complaint is fair and valid. The real world is messy and many mechanisms may not at first appear to be relevant to a particular model, but turn out to be so in the end. Thus many critics claim that simple models in fact oversimplify to the point of absurdity.

If a model ignores real world details it often becomes hard to compare such a simple model to the real world for the purposes of calibration, as well as validation. It can be difficult to calibrate a real world model

\(^2\) It is possible to have a model that uses real world data but only examines one mechanism, or a model that uses hypothetical data but examines a large variety of mechanisms and types of agents. All models are not simple or elaborated-realistic, instead there is a spectrum of models with different levels of complexity.

\(^3\) An example of the Ising model is provided as part of the NetLogo package [Wilensky, 2003]
because there is not a clear relationship between the hypothetical parameters of the model and measures that can be made in the real world. Again, in Axelrod and Hammond’s ethnocentrism model, it is not clear how to calibrate a central parameter of the model, the baseline “potential to reproduce”, to any real world data. Moreover, the output measures that can be derived from simple models often fail to have an easily understandable correspondence to the real world. Going back to Axelrod and Hammond’s ethnocentrism model, how can the frequency of an “ethnocentric genotype” be measured in the real world? Thus many critics claim that even if simple models provide general truths these truths are difficult to interpret.

The Benefits and Costs of Elaborated-Realistic Models

There are many benefits to ER models. First of all they are empirically grounded and thus there are close relationships between the parameters and mechanisms of ER models and the real world. This helps us in understanding what a ER model is telling us about the real systems we are examining. For instance, in the Anasazi model [Axtell, Epstein, Dean, Gumerman, Swedlund, Harburger, Chakravarty, Hammond, Parker, and Parker, 2003] families and water sources are represented in the places where they actually existed. Thus it is possible to use this data to test hypotheses. If the model predicts that a family would move to a particular location that has not been explored by modern archaeologists, a survey team could be sent to that site to determine if the model made a correct prediction. Moreover, even without this expenditure, it is possible to test if an ER model is correct. The ER model can be calibrated with the first half of the data and then allowed to “predict” the second half. By comparing the model’s predictions to real-world data it is possible to validate the model.

ER models allow us to explore all of the mechanisms interacting in a model simultaneously. As our models become more ER they include a larger number of mechanisms, and the interaction of this ecology of mechanisms is something that cannot be captured in any one simple model. For instance, the ER Anasazi model could have been broken up in to simple models, like a population model and a water supply model. However, only by allowing both of these different mechanisms to exist within the same context can the Anasazi model create a multi-causal account of the disappearance of the Anasazi. It is only through ER models that we can examine a holistic picture of the system we are attempting to model.

However, critics of overly elaborate and realistic models often claim that these models are not generalizable. Instead they represent particular situations and particular environments. In fact if these models are overly calibrated, they tell us nothing about the future, they simply present a reproduction of the past. This means that the model can only be used in an explanatory way. Moreover, since it is difficult to show that such a model can make accurate predictions about the future it is hard to claim that the model is validated. From a purely theoretical viewpoint, this is true. There are an infinite numbers of ways to fit a finite data set, and thus if we can not use the model to predict future data points we have no way to choose between all of the potential models. For instance, in the case of the Anasazi model it is not clear whether the model that was built correctly represents the process of the disappearance of the Anasazi or if it is simply one of many possible models. Thus as models become more elaborated and realistic it becomes increasingly difficult to generalize their results and validate their conclusions.

Moreover, since the point of modeling is to extract a “model” of the situation that is by its very nature a simplification of the world around us, a model that does not facilitate understanding is not really a model at all. Because ER models have many different mechanisms interacting in non-linear ways they are very difficult to understand and it is hard to trace a chain of causality from any class of inputs to any class of outputs. In the end if a model is elaborate enough it is better to simply observe the real world and draw conclusions about correlations rather than spending effort on a model that is not much simpler or easier to understand. Again, though the Anasazi model presents the most comprehensive description of the disappearance of the Anasazi, but it may not be applicable to many other situations. Thus, models that are no easier to understand than the real world, fail to capture useable principles that can be applied in other situations.

Full Spectrum Modeling

Full spectrum modeling combines the benefits of simple and ER modeling, while at the same time gaining additional benefits. Full spectrum modeling does not imply that there is a singular simple model and a singular ER model. Instead there can be many different simple models and many different ER models and many models at the levels of elaboration and realism in-between. By building models at multiple different levels of elaboration and realism we can extract general principles from the phenomenon we are analyzing and embed those general
principles in simple models. In addition we can calibrate ER models to real-world data and show that those models create accurate descriptions of the world around us.

We gain additional benefits from pursuing models at multiple levels of elaboration and realism at the same time. Simple models allow us to explore small numbers of mechanisms in-depth and show that those mechanisms can have profound effects on our results. We can use these simple models to determine what mechanisms should be in our ER models, and help us to validate that these mechanisms are important to the system we are studying. Vice versa, by examining ER models we can often highlight mechanisms that we want to explore in more depth and create simple exploratory models from these ER models. Thus one powerful approach to full spectrum modeling to build sequences of models either by starting with caricature-like simple models and elaborating on them, or by simplifying more elaborated realistic models.

The above approach of creating model sequences suggests a principle for the design of agent-based models. To facilitate the sequence approach to full spectrum modeling, it is important whenever possible to allow different mechanisms to be turned on and off. This allows the model to be made more or less elaborated and realistic as the need arises. The general principle of this approach to model design is to make it easy to add mechanisms and empirical data to simple models, and to make it easy to remove mechanisms and look at hypothetical data in ER models.

**Modeling Urban Formations**

Urban formation modeling has many different goals: from simply understanding how cities form to assessing the impact of interventions and different policy options on urban growth. We describe the “Cities” project, an urban formation modeling project, which we have engaged in over the past four years and in which we have utilized the full spectrum modeling approach. The goal of the “Cities” project is to create procedural formalisms of urban growth with two major purposes in mind: (1) to automatically generate urban environmental landscapes in entertainment products from movies to videogames (e.g. SimCity) [Lechner, Watson, Ren, Wilensky, Tisue & Felsen, 2006; Lechner, Watson, Tisue, Wilensky & Felsen, 2004; Lechner, Watson, Wilensky & Felsen, 2003], and (2) to build tools so that architects, interested in understanding urban processes, can understand how their conceptions and designs would influence the growth of a city [Felsen, Watson & Wilensky, 2006]. All of the models for this project have been developed in NetLogo, a freely available agent-based modeling toolkit [Wilensky, 1999].

Many attempts have been made in the past to model urban formations using agent-based models, cellular automata, and other computational modeling approaches. For instance Makse et al used a correlated percolation model [1998] and Andersson et al used a Markov Random Field model [2002]. Most of these projects used either a simple approach or an ER approach. One other project that utilized a full spectrum modeling approach to understanding urban pattern formation was the Sluce Project at the University of Michigan. This project developed simple models to explore concepts like the effect of greenbelts on urban sprawl [Brown, Page, Riolo and Rand, 2004], but also developed ER models to examine land use development patterns in Washtenaw County, Michigan [Brown, Page, Riolo, Zellner, and Rand, 2005]. For a more complete discussion of previous modeling efforts, the reader is referred to Batty [Batty, 2005].

To illustrate the full spectrum modeling approach, let us examine a progression of models that we built as part of the “Cities” project.

**A Simple Model: Path Dependence**

One simple model that we use in the Cities Project is called “Path Dependence” [Rand, Brozefsky, & Wilensky, 2006] and is based on a model by Arthur [1988]. We built this model to explore one particular thread in city development. What happens when residents in a geographic area choose their housing locations based on how many other residents are nearby? What happens when residents move to an area based strictly upon its intrinsic characteristics? What if residents make a decision that trades off between these two extremes?

In the end this model has very few mechanisms. Agents wander around a landscape and, at each time step, with a probability determined by the number of other agents located in that area and the intrinsic quality of that area, they settle in the current location. What is interesting to observe is the “path dependent” nature of the model (and hence its name). It is quite easy with proper settings of the parameters to have all of the agents settle at a suboptimal location in the search space. This model is illustrated in Figure 1.
The model we built is similar to Arthur’s model [1988] but is different in a few fundamental ways. First of all it uses a more localized search mechanism. Rather than having residents evaluate all of the locations they may possibly settle they simply evaluate locations near their current location as they wander across the landscape. Second, Arthur’s original model was based on the concept of firms and the idea that different firms would require different resources; the model we built simplifies this process by assuming that all geographic locations have one particular intrinsic quality that is important to residents.

A Second Level Model: Path Dependence with Regional Effects and Immigration

In keeping with the principles of full spectrum modeling we then created a second model that added a few new mechanisms to our original path dependence model. One assumption that we problematized is that it seems unrealistic that individuals are only concerned about the immediate neighborhood that they settle in. Instead they can also be concerned about other neighborhoods nearby. To assess this issue we added regional effects to the path dependence model. Regional effects were taken into account by having the residents examine the number of residents and intrinsic quality of nearby neighborhoods when deciding whether or not to settle in the current location. A second issue with the simple path dependence model is that there is no mechanism for immigration. After all city populations do not remain constant in time. Thus we added immigration to the model by adding a new parameter that generates new agents every time step. This results in a non-constant population with new residents, potentially with new preferences entering the model over time. The neighborhood effect means that settlement patterns are more dispersed across the entire world since desirable areas are larger. The immigration effect means that the areas with the highest population can change any time and the model no longer simply predicts equilibrium. Thus adding mechanisms to the model has enabled us to produce more realistic results. We have developed many other intermediate models of urban formation but we now advance to a very elaborate and realistic model of city development.

An Elaborated-Realistic Model: PMC

The most elaborated and realistic model that we have developed as part of the Cities project is the PMC model. This model involves many different agent types (developers, residents, business owners, policymakers, highway developers, infrastructural agents, and many more) and has many different land-use types (resident, commercial, industrial, water, hill, park, road, and others). Moreover the model user has control over a large variety of parameters. They can seed areas of the environment as being more conducive to development by particular developer types (commercial, residential, and commercial) and they can physically alter the world to place water and elevation in certain areas. They can even control where roads will be “gridded” and where they will be more naturally formed. Adding all of these controls and mechanisms to the model has allowed the creation of very realistic output. In fact this model has been used by classes of urban architects to create models of real cities like Berlin, Jerusalem, and Paris. The results from a typical run of the model are illustrated in Figure 2.

Summary of Cities Project Experience

By building these three different models (and several others that we have built for the “Cities” project) we have learned a lot about both city formation and about how to do full spectrum modeling. One important design principle is the importance of building models so that they can be simplified or enriched easily. For instance, when the PMC model was first created we built it with the assumption that to create a model of a city all of the important features and agents would need to be constructed, and we did not provide for the possibility of simplifying the model by eliminating mechanisms or agents. However as the project has progressed, difficulties with examining all of these layers have led us to spend effort removing some of these layers and simplifying the so that we can run the model with fewer agent types and mechanisms. This effort is making it easier for us to examine the results and to extract some of the mechanisms for which we will create simple models. This process has been facilitated by the creation of our path dependence models. By examining how these simple models work and by showing that they create

4 In fact the implementation of these two models is in the same code base. Though they represent different conceptual models and levels of complexity they can be represented using one source code file.

5 PMC stands for “Procedural Modeling of Cities.”
results similar to the more ER mechanisms in the PMC model, we have been able to remove some of the elaboration and realism in the PMC model and still achieve comparable results.

**Conclusion**

We have argued that full spectrum modeling enables a richer understanding of complex phenomena. By examining systems at both a simple level, which gains us insight into general principles, and at an elaborated and realistic level, which allows us to understand the model’s ramifications on the real world, and at multiple levels in-between, we gain a richer and deeper understanding of the phenomenon we are attempting to model. We hope to continue this work in general by examining how each of these levels of models aids our understanding of complex systems. We also hope to continue this work in particular by building more and more models of urban pattern formation at various levels of detail. In the end, there is no principle of modeling that tells us exactly at what level of detail to build a model at to best understand the world around us, instead the solution is to build multiple models and learn from each of those models independently. By comparing, elaborating and refining models at multiple levels and building progressions of models we learn much more than it is possible to learn from any one model.

**Acknowledgements:** We would like to thank Ben Watson and Martin Felsen for their collaboration on the “Cities” project and the development of the PMC model. Tom Lechner and Seth Tisue also greatly contributed to the PMC model development. We would also like to acknowledge the assistance of Craig Brozefsky both in developing the PMC and Path Dependence models, and also for his thoughts and insights into the full spectrum modeling approach.
References


Figure 1: Path Dependence model in NetLogo based on Arthur's Model (http://ccl.northwestern.edu/cities/)

Figure 2: PMC Model in NetLogo (http://ccl.northwestern.edu/cities/)