Copyright
by
Scott Austen Robinson
2014

The thesis committee for Scott Austen Robinson
certifies that this is the approved version of the following thesis:

The Solar Energy Consumer Agent Decision (SECAD) Model: Addressing complexity through GIS-integrated agent-based modeling

APPROVED BY

SUPERVISING COMMITTEE:

Varun Rai, Supervisor

Eugenio Arima

Jay Zarnikau

The Solar Energy Consumer Agent Decision (SECAD) Model: Addressing complexity through GIS-integrated agent-based modeling

by

Scott Austen Robinson, B.A.

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degrees of

MASTER OF ARTS

AND

MASTER OF PUBLIC AFFAIRS

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2014

This work is dedicated to my parents, Will Robinson and Maria Katherman
for showing me the value of curiosity.

Acknowledgments

I wish to acknowledge the hard work of others in the Energy Systems Transformation Group that made this work possible. Without years worth of survey and program data we would not be in a position to run simulations at this scale and level of detail. I would like to acknowledge the tireless support and guidance of Dr. Varun Rai, whose vision and understanding of the underlying issues was irreplaceable. I would like to thank the National Renewable Energy Laboratory and the United States Department of Energy for their support and funding of this research, and Austin Energy, Frontier Associates, the City of Austin, and Travis County for providing data. Finally I would like to acknowledge the Texas Advanced Computing Center for proving the computing power that made city-wide agent simulation possible.

The Solar Energy Consumer Agent Decision (SECAD) Model: Addressing complexity through GIS-integrated agent-based modeling

Scott Austen Robinson, M.A.; M.P.Aff.
The University of Texas at Austin, 2014

Supervisor: Varun Rai

This thesis presents a step-by-step implementation of the Solar Energy Consumer Agent Decision (SECAD) model: an empirically-grounded multi-agent model of residential solar photovoltaic (PV) adoption with an integrated geospatial topology. Solar PV diffusion is a complex system with geographic heterogeneity, uncertain information, high financial risk, and important social interaction and feedback effects between consumers. A key limitation for agent-based models in human socio-technical systems is the integration of empirical patterns in the model structure, initialization, and validation efforts. This limitation is addressed though highly granular and interlocking data-streams from the geographic, social network, financial, demographic, and decision-making process of real households in the study. The fitted and validation model is used to simulate implementation of potential policies to inform decision-makers: i) Targeted informational dissemination campaigns, ii) Tiered rebates, iii) Locational pricing, and iv) Alternative rebate schedules. Informational campaigns

can increase cumulative installations by as much as 12%, but vary greatly in their effectiveness based on which agents are targeted. Simulations suggest that by lowering the cost barrier to lower wealth households through a slightly higher rebate (+$0.25/Watt), the mean difference in wealth between solar adopters and non adopters could be reduced by 22.6%. Locational pricing can allow the utility more control over diffusion patterns with regard to load pockets–a $0.25 higher offering increased the percentage of adopters in the target area from less than 1% to over 10%. Relative to flatter rebate schedules, sharply decreasing schedules are effective in terms of motivating adoption but inefficient in small markets. It is our hope that this work will provide a working example for other agent-based models of human socio-technical systems as well as provide insight into the likely outcomes of novel policy-levers such as those described above.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Diffusion of Technology

The rate at which new technologies are incorporated into an economy has long been linked with the performance of that economy, and is nearly inseparable from the idea of competitiveness on an international scale [45, 105]. The process by which this occurs is generally referred to as diffusion, in reference to the movement of the technology through the population over time [48]. Because the diffusion process is so tightly bound to economic competitiveness, and examples of successful and unsuccessful innovations are abundant, the literature pertaining to diffusion of technologies has flourished in economics, [12, 55, 140], marketing [47, 93, 116], sociology [21, 41, 127], and policy analysis [70, 71, 100]. In order to predict technology diffusion outcomes and whatever benefits or costs might be associated with that diffusion, it is necessary to model the diffusion process.

Fundamental to diffusion of innovations theory is the idea that technology diffusion is quantified as the cumulative sum of the adoption states of individuals. In other words, what researchers measure as diffusion is the aggregate result of the individual decision-making process for each consumer. Therefore, successful models will simulate this individual process, at least on the aggregate. While aspects of consumer decision-making have been modeled with varying degrees of accuracy and abstraction using econometric models [71, 133], stochastic models [13, 51], and Geographic Information Systems (GIS) [76, 84], three theoretical models are commonly used to explain product diffusion as an aggregated decision-making process: the consumer heterogeneity or Bass model, the contagion model, and the real options model [61, 93].

1.2 Diffusion Modeling in Socio-Technical Systems

1. The consumer heterogeneity (Bass) model. This model starts from the reasonable assumption that consumers vary in the costs and benefits they associate with a technology. These differences are attributed to heterogeneity in tolerance for risk, values, or information. For example one person might value staying on the cutting edge of technological change, while another might place more value reliability or proven functionality [127]. The most common implementation of the consumer heterogeneity model is the Bass model [10], where marginal adoption S at a given time t is a function of the size of the market m, time, and two coefficients p and q representing advertising and word of mouth effects:

$$S(t) = m\frac{(p+q)^2}{p}\frac{e^{-(p+q)t}}{1+(q/p)e^{-(p+q)t2}} \tag{1.1}$$

Because Bass models only employ consumer heterogeneity along one dimension, the 'innovativeness' of the individual relative to the rest of the population, there is little potential to incorporate attributes that have been demonstrated to affect consumer purchase decisions, such as the availability of capital. Further, the model makes it difficult to incorporate price shocks, multiple influencing institutions, or competing products, and can incorporate spatial relationships only on the aggregate, for example through multiple state-level specifications. While some econometric methods such as accelerated time to failure or proportional hazard models allow for the handling of multiple attributes, in each case heterogeneity is reduced to aggregate terms, and the inclusion of network effects, institutional influences, and uncertainty remain problematic [123].

2. The contagion model. In this formulation, costs and benefits to consumers are essentially held constant, and individuals make adoption decisions as they gain information about the technology. Information passes from one agent to another through proximity in a network. Consumers become informed (active) after another consumer in their network does. Once active, consumers are infectious, meaning they spread information to their network connections in the form of an adoption probability. Over time, more consumers are informed and adopt the technology. As the population reaches saturation, marginal adoptions continue but the rate of adoption decreases [141], creating an S-shaped curve. One of the most common implementations of this model is the cascade model [79], which in a simple form where b is the number of contacts for each consumer, and ϕ is the ratio of non-adopters to adopters at time 0, results in a

logistic function [48]:

$$S(t) = m\frac{1}{1 + \phi e^{-pmt}} \tag{1.2}$$

Some authors have categorized consumer heterogeneity models as a type of contagion model with internal and external influences [153]. In practice, contagion models often employ a threshold (number of infected connections) in order to fit diffusion rates to empirical scenarios [145], and the model can incorporate spatial relationships implicitly through the initialization of the network. Many current applications of these models utilize social network graphs [88]. The biggest problem with these models lies in the simplicity of their assumptions. Contagion models do not model consumer heterogeneity outside the attributes of the social network–two nodes placed identically in the network will always adopt at the same time or with the same probability. In other words, the network attributes of the nodes, such as degree, centrality, and density (see section 1.6) correlate perfectly with underlying diffusion determinants. This assumption is difficult or impossible to justify empirically or theoretically.

3. The real options model. This model is seen in the economics literature as an improvement on deterministic discounted cash flow approaches [82], mainly through recognition of the economic value of waiting (comparing the net-present value (NPV) of an investment at time t to the NPV of the investment at time $t + 1$, expressed in equation 1.3), where u_t is an investment opportunity yielding some profit π. The continuation value

(expectation of net present value at time $t+1$) is discounted by the factor $\frac{1}{1+\rho}$.

$$E_t(NPV) = \max_{u_t}(\pi_t(u_t) + \frac{1}{1+\rho}E_t(NPV_{t+1})) \tag{1.3}$$

Here the adoption decision is cast as an investment decision under uncertainty, and although it is not explicit, the value calculation can be viewed as a decision threshold for the consumer or firm. An adopter faces an upfront investment that is treated as a sunk cost, to be offset with a future returns stream. The consumer is generally thought to have accurate information regarding investment costs, and uncertain information regarding future benefit. Rather that being derived from consumer attributes, uncertainty is modeled by a stochastic process. Importantly, the consumer can exercise the buy option at a future time, when the cost-benefit calculation may be more favorable. This creates incentives for delay in adoption of products with declining costs, uncertain payoffs, and low "reversibility" (the ability to un-adopt cheaply) [11, 89]. Real options models seem to have great potential, although applications to empirical diffusion scenarios are fairly scarce [78], and require the assumption (and choice) of set discount and learning rates [82]. They also do not have social network information sharing or spatial relationships modeled explicitly, and because the driving component of the model relies on an NPV calculation, consumer rationality is assumed.

In part because of the problems noted above with the consumer heterogeneity, contagion, and real options models, applying of traditional models to diffusion of innovations in socio-technical systems remains problematic. Weak points of traditional diffusion modeling are compiled in a 2004 review by Kemp [78]. Seven of these are particularly relevant to socio-technical systems and are summarized below:

1. Focus on the macro social process rather than modeling decisions explicitly.
2. The full incorporation of social and network effects.
3. Lack of feedback between knowledge transfer and changing economics.
4. Failure to incorporate the actions and effects of policy and policy actors.
5. Failure to address uncertainty in the *time of adoption* by a given consumer.
6. Lack of consideration for competing technologies.
7. Failure to incorporate complementary innovations and infrastructure growth.

The unifying theme of this criticism is that diffusion models suffer from a failure to address the complexity of human systems. While the ease of implementation of these models makes them attractive, this simplicity also makes the above models limited in their explanatory power. If we model a system to gain useful insight into how it operates and how it might be altered, an alternative method must be employed.

1.3 Agent-Based Modeling

An alternative option, agent-based modeling (ABM), is flexible enough to incorporate the strengths or each of these models, while avoiding weaknesses, such as aggregation, handling spatial relationships, modeling heterogeneity across multiple dimensions, feedback effects, and emergent phenomena. ABM is a simulation technique in which the entities in a system are represented as autonomous functions that have memory and interact with each other and their environment, which is modeled explicitly [90, 108]. Relative to competing methods, the strength of ABM is derived from the ability to capture the heterogeneity of individuals and interaction effects in the agent class [29, 117]. By explicitly coding the decision rules for agents, the processes that lead to aggregate results can be probed and altered experimentally [4, 29, 90]. Allowing emergent processes to be modeled is particularly useful in complex systems as outcomes are often counter intuitive [17].

1.3.1 State of the Practice

While the potential of ABM to address the weaknesses of conventional diffusion models in socio-technical systems is very promising, and state-of-the-art models can do so, these are few and far between. Current ABM practices as applied to human systems are lacking, mostly due to inadequate validation [12, 35, 76, 149], or inadequate empirical basis [16, 25, 35, 36, 146].

Though not always done in practice, the need for empirical basis and validation in the model has been recognized [56, 122]. By tying the agent states, decision rules and environmental variable to empirical patterns, ABM gains descriptive, explanatory, and predictive power. Seven principles of empirical ABM are presented in detail in chapter 2.

1.3.2 ABM and GIS

As noted in section 1.2, traditional models do a poor job of incorporating spatial relationships. This is fairly simple in ABM through specification of the topology and environment [108]. This coupling of ABM and Geographic Information Systems (GIS) can increase the empirical basis of the model (figure 1.1), and allow greater precision in agent attributes. While increased precision can have large computational costs, leading many to employ aggregated agents, or stylized environments and topologies, the increasing accessibility of high performance computing (HPC) resources makes these costs less significant.

Specific to solar PV, ABM integrated with GIS offers four distinct advantages, all of which contribute to the ability of the ABM to generate realistic outcomes during simulated scenarios that have not been directly observed in the study area [117]. First, each household in the model can take on the actual attributes and street location of the real household it represents. Because neighborhood peer effects have been demonstrated to be the most relevant social interactions for solar PV diffusion [118], it is much more likely that our agent-to-agent interactions simulate the actual relationships in the study area. Second, there is the possibility to examine and utilize derived network attributes, such as the

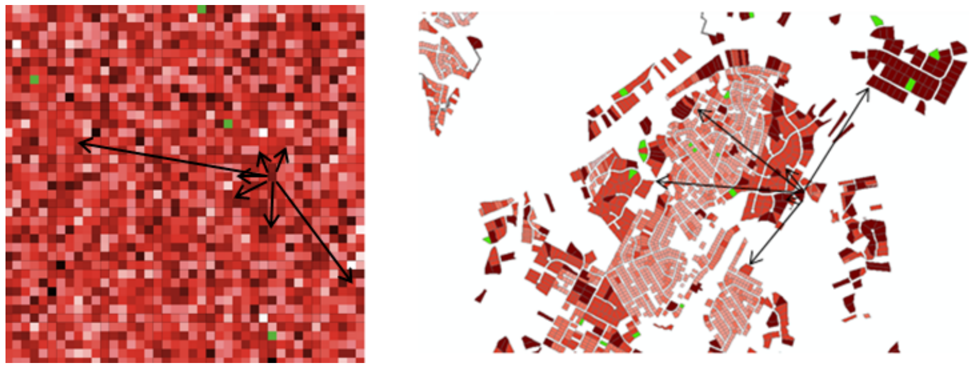


Figure 1.1: Grid versus GIS topology in the SECAD model.

location of influential households, which can be utilized in policy simulation scenarios. Third, it becomes possible to simulate at an extremely granular level not only when, but where PV will come on-line, allowing policymakers to plan and target infrastructure investments, and see the geographic implications of different policies. Fourth, GIS integrated ABM allows us to delve into important issues of social demographics and equity associated with PV that may be exacerbated by peer effects and local information flows.

Before moving on to the discussion of the principals of empirical ABM, it is useful to provide a more complete picture of consumer decision-making from a social/behavioral standpoint. While limited, the three diffusion models discussed above have demonstrated the necessary aspects of a robust diffusion model of a socio-technical system: heterogeneity, social networks, and uncertainty. The remaining sections of this chapter will be dedicated to further detailing these issues towards their application to a robust ABM of solar PV diffusion.

1.4 Theory of Planned Behavior

The fundamental insight of heterogeneity models is that consumers vary in their perception of value, which has a direct influence on behavior. Perhaps both the most nuanced and validated behavioral model that can be readily incorporated into a computational framework comes from the theory of planned behavior (TPB) [6, 52], which itself arises from the theory of reasoned action [2, 92]. TPB starts with the assertion that any action is the result of an intention to perform that action, limited by one's ability to perform that action. Intention is a function of an individual's attitude (his/her overall assessment of the probable outcome), subjective norm (the beliefs and information available through social connections), and perceived behavioral control (the consumer's perception of his/her own ability to perform the action) [3].

In an applied context, the three components of intention can be modeled as a function of attitudinal, social, and demographic variables [94, 103]. This framework has been successfully applied in a number of ABMs, for example: theoretical markets [154], human migration [81], dietary choice [125] and the diffusion of organic farming [76], smart meters[155], pellet burning stoves [138], and water saving innovations [130]. While the importance of modeling agent beliefs is gaining recognition, many models still avoid or ignore this aspect, or represent it in an over-simplified way [142].

1.5 Information and Uncertainty

The incorporation of uncertainty in information in the real options model is an important step toward a more realistic view of the way costs are evaluated. The true cost of a technology to consumers is often much above the hardware or sticker price given by the supplier. Uncertainty in the cost-benefit associated with the technology and its alternatives takes time and energy to resolve. The potential for error or missed opportunity creates a disincentive for the decision to be made quickly, leading to the information search process and its associated costs[80, 104, 139]. As each consumer takes a longer time to make a decision, and perceives the decision as being more risky, the rate of diffusion decreases [61].

The adoption decision requires multiple information inputs. First there is the knowledge that the product exists. Second is the suitability of the product to the consumer's needs. This information comes to the consumer from two main paths: from producers of the technology, and from interactions with other users, often through geographic proximity [61]. In the presence of informational externalities, these interactions can be part of a social learning process used to address the uncertainty associated with an innovation: potential adopters wait and observe the experience of others [74]. This perspective is emphasized in the diffusion of innovations literature in sociology [127]). For nascent technologies, because the social network is relatively unsaturated with information, the amount of time and effort spent by the consumer is greater. This effect is further amplified for capital intensive technologies because the potential down-side is greater [89, 115]. Clearly, imperfect information leads to the inefficient operation of markets as consumers deal with uncertainty, creating the potential for market failure [57, 128]. For technologies with significant positive environmental externalities a compounded market failure can be created as the external benefit associated with the environmental good is combined with the external cost of information search. Together this can cause a large barrier to diffusion [72].

1.6 Social Network Analysis

The implicit (consumer heterogeneity model) and explicit (contagion model) assumption that diffusion of technology and information occurs through interpersonal connections, as well as the importance placed on social systems in diffusion of innovations theory [127] motivates a clearer understanding of social network structure and theory. To quote E. Katz, "It is as unthinkable to study diffusion without some knowledge of the social structures in which potential adopters are located as it is to study blood circulation without adequate knowledge of the structure of veins and arteries [75]." In this section, the focus will be on two main topics: social network metrics and social network models. Social networks are represented in scientific models by graphs on nodes connected by edges, usually stored in an adjacency matrix of binary values. Each node represents one person or households, and each edge represents a connection or relationship between two nodes.

1.6.1 Social Network Metrics

Social network metrics are numeric values used to characterize nodes within a graph, or the entire distribution of nodes [132].

1. **Degree** The number of edges incident on a node. The social network can be characterized by the degree distribution. If a graph is directed, meaning that the relationships in the network are not symmetrical, degree can be broken down further into in-degree and out-degree. Degree can be thought of as a measure of centrality for a given node, although it is quite possible for a node with high degree to have low centrality measured in a different way (see below).

   In Degree Number of other nodes with edges terminating in node i. Shown for adjacency matrix A in equation 1.4:

   $$D_{Ii} = \sum_{i=1}^{n} A_{ij} \tag{1.4}$$

   Out Degree Number of edges with origins in node i. Shown for adjacency matrix A in equation 1.5:

   $$D_{Oi} = \sum_{j=1}^{n} A_{ij} \tag{1.5}$$

2. **Centrality** The importance of a given node can be measured in different ways. The relevance of given measure will depend on the network being characterized and the problem the researcher hopes to address. For example, in a corporate hierarchy more important nodes may not be

those that are near the center of the graph, but rather those that act as brokers between different groups.

Betweenness The degree to which node i acts as a broker between different groups, due to that node lying on the shortest paths g_{jk} between any other two nodes j and k:

$$\frac{C_{Bi} = \sum_{j<k}^{n} \frac{g_{jki}}{g_{jk}}}{(n-1)(n-2)/2} \tag{1.6}$$

Closeness The degree to which the rest of the network is accessible from node i, also thought of as the shortest number of steps or distance d on the graph (in edges) from i to all other nodes j. This is shown in equation 1.7.

$$C_{ci} = \frac{(\sum_{j=1}^{n} d_{i,j})^{-1}}{n-1} \tag{1.7}$$

Eigenvector centrality Intuitively, the importance of node i in the network is influenced by the importance of i's neighbors on the graph. This is not fully captured by the above metrics, but can be calculated iteratively (as the centrality of a node is calculated, the centrality of its neighbors must be updated) though equation 1.8 [18]. It has been noted that Eigenvector centrality is uniquely suited to calculate centrality for complex and irregular graphs, particularly those with non-binary relations between nodes [19]. The coefficient β can be varied to reflect the importance of neighbor's centrality relative to the centrality of i. As β approaches zero, centrality reflects more of the global network structure. α is a normalization constant. If Eigenvector centrality is calculated on a directed network, this is known as PageRank [110].

$$C_{Ei}(\beta, \alpha) = \sum_{j=1}^{n} (\alpha + \beta C_{Ej}) A_{ij} \tag{1.8}$$

Centralization Some networks are more highly centralized than others. The amount a variation in the distribution of centrality scores for the nodes is known as centralization. There are multiple metrics, including standard deviation, the gini coefficient [87], or Freeman's centralization, which measures how central the most central node is relative to the other nodes on the graph (C_x), where $C_x(p_i)$ is any centrality measure (see above) for a given node p_i, and p_* is the most central node (shown in equation 1.9) [46].

$$C_x = \frac{\sum_{i=1}^{n} C_x(p_*) - C_x(p_i)}{(N-1)(N-2)} \tag{1.9}$$

3. **Connectivity** Components in the network are strongly connected when each node can be reached through directed links by any other node. Components in the network are weakly connected when each node can be reached through undirected links by any other node. If the largest component takes up a significant portion of the network, it is known as the giant component.

4. **Average Shortest Path** Mean number of steps or distance ($d_{i,j}$) needed to move along edges from one randomly chosen node to another.

5. **Local Clustering** Also called transitivity: the number of node i's connections that are connected by an edge, ν_i, over the total number of potential connections [150]. Transitivity can thus be thought of as the probability that adjacent nodes are connected. This is shown in equation 1.10 for a directed network. For diffusion of innovations studies, higher clustering is associated with slower information flow compared to random graphs [109]. This slightly counter-intuitive result is caused by the lack of "short-cuts" in highly localized networks.

$$Cl_i = \frac{\nu_i}{n_i(n_i - 1)} \tag{1.10}$$

Equation 1.11 shows the global clustering coefficient as simply the average of the local clustering values Cl_i over all nodes n.

$$Cl = \frac{1}{n} \sum_{i=1}^{n} Cl_i \tag{1.11}$$

1.6.2 Social Network Models

Social network models are simplified representations of empirical networks, used to derive and predict properties and outcomes of the network mathematically [23, 132]. The intent of this section is not to provide the reader with a summary of all network models, but rather to provide details of several models that will be useful in developing an understanding of how the model specification might effect social network metrics and diffusion.

1. **Erdos-Renyi** This undirected network model assumes that nodes are connected at random with some probability p. As the individual connections are independent Bernoulli variables, the degree distribution is described by a binomial distribution with parameter p.

2. **Static Geographical** This network model connects each node to a number $NumC$ of its closest neighbors, which are defined by proximity in space. Alternatively, each node can be connected to each of its neighbors within a set distance r. Relative to Erdos Renyi graphs, Static Geographical networks have longer average shortest-path length. If parameterized by $NumC$, the degree distribution will tend to be tighter, while if parameterized by r the degree distribution will reflect the density of nodes in space. These networks often display structure similar to that of a regular or semi-regular lattice.

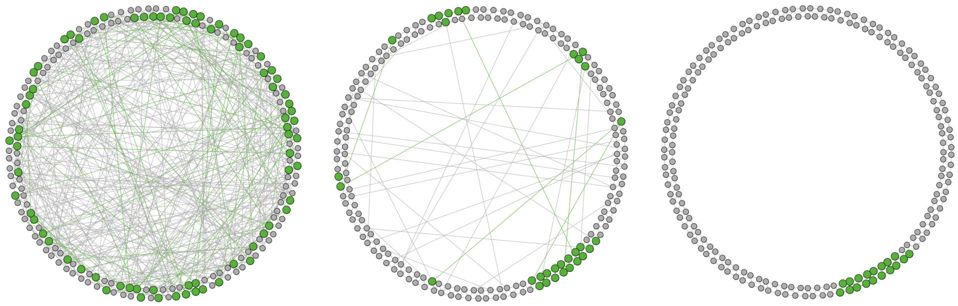
3. **Small-World** This model is characterized by high clustering relative to Erdos-Renyi random graphs, and low average shortest-path. These networks have been observed in a variety of empirical settings [1, 102, 129, 150]. Small-word networks can be conceptualized in a modeling setting by initializing a static geographical or a lattice network, and re-wiring edges on the graph with some probability p, holding the degree

constant. This achieves a network in which the majority $(n - n * p)$ of connections are local, and the minority $(n * p)$ are non-local, or random.

The structure of the network can have a large impact on diffusion outcomes, making the specification of the network critical [113]. This been well-studied in simulated networks, and positive effects result from word-of-mouth and rewiring, and negative effects result from network externalities where an innovation becomes more useful when more nodes have acquired it [53].

To illustrate this point, a simple model was created in Netlogo to illustrate the diffusion outcome for simple networks each made of 200 nodes, each with four edges. The model is shown in figure 1.2. In each case, one node was seeded with an innovation (shown in green), and the model was advanced ten time-steps. At each time-step, the innovation was given a probability of 0.25 of spreading through an edge to a connected node. Under these simplistic conditions, on average the random graph (a) displays more rapid diffusion than the small-world graph (b), which displays more diffusion than the static geographical graph (c). These findings support previous conclusions from the literature.

One limitation that becomes clear with this type of model is the way that the consumers (nodes) make the decision to adopt. In this simplified model, the decision is made as a probabilistic outcome of contact. While this may be adequate for some disease or for theoretical models [15, 83], a more nuanced framework is necessary in order to gain the benefits of social network analysis



(a) Erdos Renyi (b) Small-World (c) Static Geographical

Figure 1.2: The effect of Network structure on diffusion outcomes.

in an empirical diffusion model.

Chapter 2

Principles of Empirical Agent-Based Modeling

2.1 Motivation

The expanding popularity of agent-based models (ABMs) in the characterization of human systems (for example in economics [124], psychology [137], and anthropology [144]) demands further development and unification of ABM methodology to allow researchers to gain the full benefit of this technique. While ABM can be used for virtually any modeling application [40], the strengths of ABM are most effectively utilized in the exploration of previously unobserved situations, or scenario analysis [9]. Consequently ABM allows for the creation of virtual laboratories used in decision support roles. However, the usefulness of this analysis will depend on the degree to which the ABM reflects real-world conditions. While this is intuitive, too often this is overlooked or ignored.

The focus on socio-technical energy systems, and particularly the diffusion of environmentally friendly technologies, is driven by need. Cumulative diffusion trends for a given technology are driven by decision-making and interactions within a heterogeneous population of consumers, which traditional modeling frameworks do not adequately reflect [61]. As a further complication, these interactions, the attributes of the consumers, and the characteristics of the technology depend on the specific geographic context [67, 120, 147]. While there have been many applications of ABM in the diffusion of technologies field in recent years (for example see [22, 58, 85, 131, 143, 149, 155]) the success of these models has so far been impeded by the lack of empirical foundation, especially in regards to the geographic relationships of the agents and environment.

While the ABM discipline as a whole has begun to shift toward empirical models [134] following previous calls for agent-based modeling to increase its focus on validation [56, 108, 122], these efforts have so far gained little traction within human systems. This is likely due to the complexity of these systems and the level of data required (there are a few notable exemptions: for example [42, 64, 86]).

In order to guide the development of more useful agent-based models for environmentally-friendly technologies, seven principles are proposed that will help researchers move away from proof-of-concept or toy models (for example see [68]) and toward decision-support models. To ground each principle, concrete examples are given to illustrate each principle. The seven principles are the following:

1. Empirical Agents
2. Theoretical Validation
3. Networks and Interaction

4. Empirical Environment
5. Precise Initialization
6. Empirical Validation
7. Simulation Process

The central thread tying these principles together is the use of empirical data bring the model closer to real-world conditions. While this thesis is specifically intended to present the SECAD model of solar PV diffusion, several other socio-technical ABMs are discussed to provide a deeper context for empirical ABM principles.

2.2 Empirical Agents

The empirical agents principle has three components: use of data, level of aggregation, and agent classes. It is worth noting that the data required to design a successful ABM is rarely collected in one round. Rather, the level of aggregation in the proposed model can drive data collection. Analysis of the data provides a basis for new agent classes and validates the proposed level of aggregation, which in turn motivate collection of new data. This feedback is vital to the creation of robust models.

2.2.1 Use of Data

Using real-world data to design and validate the ABM and move away from toy models builds on other authors' discussion of ABM parameterization in human systems and the use of empirical data [43, 136]. Perhaps more than any other modeling technique, the danger of drawing erroneous conclusions from an invalid assumptions ("garbage in, garbage out") is prevalent in ABM. To avoid this, modeling must be preceded by deep micro-data collection and analysis, perhaps through multiple, integrated, and iterative levels. Examples of this is the use of household level survey data with program data [126], and combining survey data, expert interviews, and census data [135]. Methods for collection of sample data and up-scaling the data to population parameters (or down-scaling from aggregate data) have been thoroughly addressed by other authors [8, 135, 136]. Initial models can provide direction to further data collection, and step-wise refinement.

2.2.2 Level of Aggregation

The level of aggregation defines the unit at which agents operate. It can mean moving from modeling individuals to modeling groups of individuals. While it is tempting to base aggregation on the availability of data, this is not appropriate. The level of aggregation chosen for a successful agent-based model must allow the ABM to be motivated by the same mechanisms that drive the target system. If the agents are so abstracted that the mechanism must deviate from that observed empirically, more granular agents should be

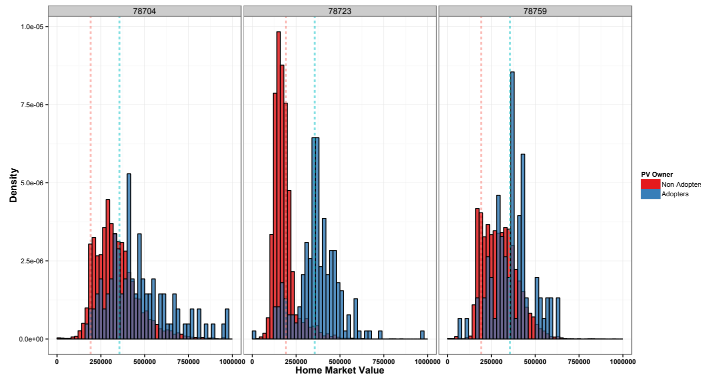


Figure 2.1: Wealth distributions for PV owners versus population wealth distributions in three zip codes.

used. For example, the diffusion of residential solar PV is driven by individual consumer decision-making, while the diffusion of wind turbine generators is driven by firm-level decision-making. Using zip code level agents for solar PV is clearly too aggregated, while individual level agents for wind is too granular; the use of each would require a fundamental departure from the target mechanism.

To elaborate on the PV example above, the distribution of adopter attributes may not reflect zip code level attributes. This is evident in density distributions such as the one displayed in Figure 2.1. In the figure, the distributions of wealth across four zip codes among adopters are distinct from population distributions. This demonstrates the need for the model to operate below the zip code level. The mechanisms behind successful ABMs will try to reproduce as closely as possible the way the target system operates.[156]

It is important to note that a successful ABM can contain multiple levels

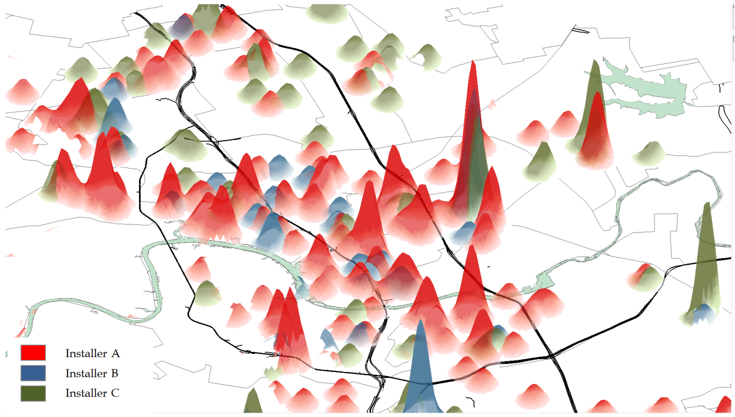


Figure 2.2: Local marketing effects visible in the distribution of PV systems over space.

of aggregation. Within the wind model for example, firm-level actions could coincide with state-level policy modeling. The different levels of aggregation exist for different agent classes.

2.2.3 Agent Classes

The complexity of the systems that are best suited for ABM often require additional agents to increase the realism of the model. This is appropriate when there is a quantifiable impact on the target output that is driven by an autonomous, heterogeneous force. In the case of solar PV, the solar installation companies may pass this test. While individual consumers are the primary agents within a solar PV model, the installer-specific density of empirical data shown in figure 2.2 suggest autonomy and heterogeneity that justify creation of an agent class to represent solar installers.

2.3 Theoretical Validation

Theoretical validation of an ABM means assessing the degree to which the model is an accurate representation of the target system, as defined by existing theory and empirical data, with reference to the central research question.

2.3.1 Existing theory

Existing theory does three things: it places the model within its scientific context, it guides the model structure, and it supports the choice of algorithms to codify decision rules.

Because ABM is a simulation method, the model must be derived from existing theory in the scientific literature of the system. Otherwise, the functions (behavioral rules) that define agent behavior will be arbitrary. Arbitrary rules will not generate model results that can easily contribute to scientific understanding of the target system.

While the emphasis of this work is on the use of empirical data to provide the basis for the ABM, neither raw data, nor supporting models will provide model structure. In a simple example, empirical data shows that wind turbine hub heights have been increasing over time, but does not provide a mechanism for this increase. How are the agent's actions driving this pattern? The use of theory to guide the gathering, analysis, and incorporation of data into the model is critical to an effective approach [43].

There are potentially many algorithms that can be written for a given theory, and many theories that describe a target system. Using an example from human behavior, agents can be programmed according to CODA (continuous opinions, discrete actions) for opinion dynamics, rational choice, or theory of planned behavior. In this case, one useful application of ABM is to create batch parameters (see section 2.8) or scripts that simulate the target system according to each theory. The batches can then be evaluated for accuracy against the empirical data, as shown in section 2.7.

2.3.2 Empirical patterns and decision Rules

Like most statistical and computational models, highly effective ABMs require a thorough understanding of the processes that generated the data and of the critical patterns in the data that motivate the research question. The model structure should reflect descriptive findings from empirical data uncovered during data exploration. This concept is related to what has been called Patten Oriented Modeling or POM [56]. The idea is to shape the structure of the model based on multiple patterns observed in the target system, where patterns generally reflect data-driven relationships and correlations. This idea has been applied in the modeling of biological systems [5, 98, 101], but has not yet caught on for human processes. Successful agent-based models validate the theoretical basis for parameter choices and decision rules with reference to empirical patterns. For example, analysis of data on solar and electric vehicle purchases suggests that the technologies have different social network structures. Figure 2.3 shows that the networks that define solar PV diffusion

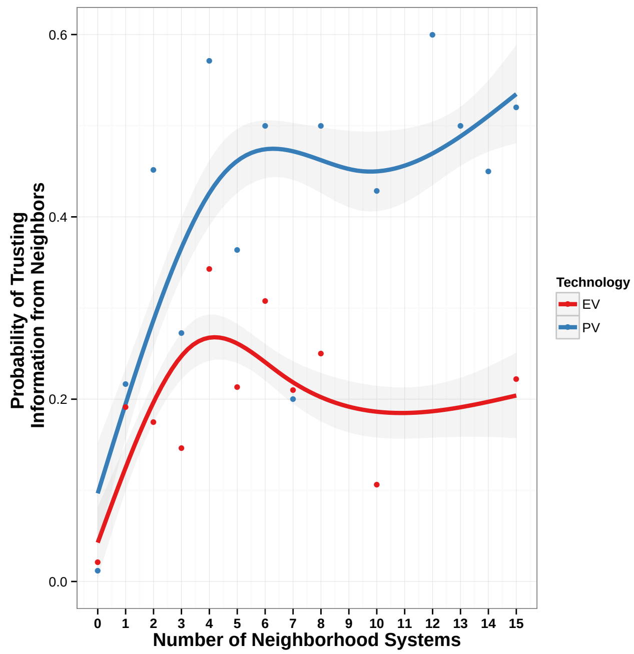


Figure 2.3: Trust and local network structure in solar PV versus electric vehicles.

are driven by geographically local connections. This is much less true for electric vehicle diffusion, which is driven by higher dimension networks, as discussed in section 2.4

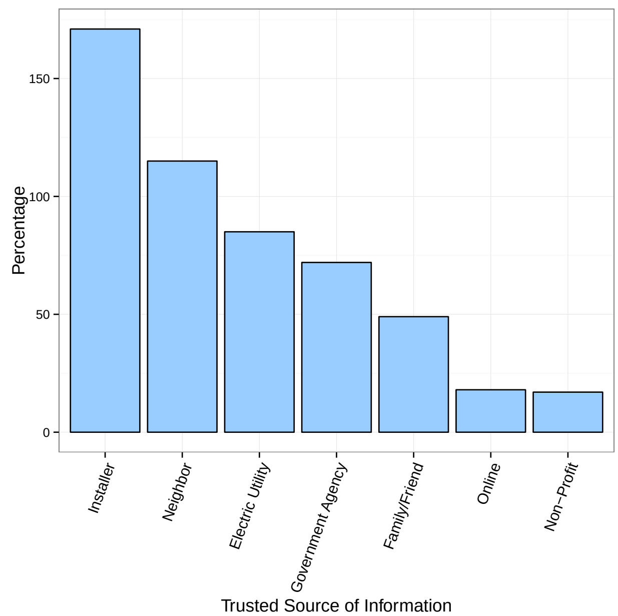
Without knowledge of this empirical pattern, it could be quite plausible to specify the same network structure for both of these technologies and end up with an invalid model for the system.

2.4 Networks and Interaction

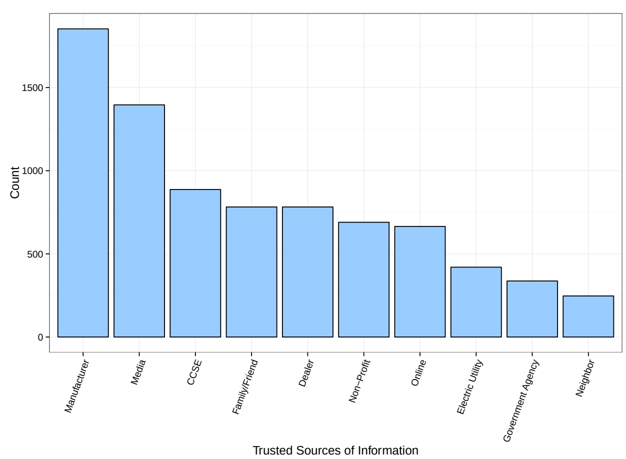
While nearly all agent-based models of socio-technical systems incorporate social networks, the most common practice is to position the agents randomly in a network structure [16, 35]. This can have profound effects, as adoption outcomes are strongly influenced by proximity to central nodes in the network, and the state of those nodes. If network characteristics are assigned randomly, these effects will either remain external to the model if a new distribution is created for each simulation run, or they will bias the model outcomes if one distribution is used for multiple simulation runs.

The appropriate dimensionality of a network is dependent on the technology or process being modeled. Dimensionality is the number of levels on which nodes must be empirically derived. For example, Figure 2.4 shows that the sources of information used by potential PV owners during their research periods are much fewer, and more concentrated than those used by potential EV owners. This means that robust networks for PV will require fewer specified dimensions than EV networks. There is strong evidence that for solar PV, the most influential nodes are those of geographic neighbors [118] and installers [120]. In electric vehicles, by contrast, these connections are less important, while those with manufacturers, media, friends and family, and others are more important.

If there is interaction between agents in the model, the network definitions must be specified with care. Several of the problems highlighted in Section 2.3



(a) Solar PV



(b) Electric Vehicles

Figure 2.4: Trusted sources of information in electric vehicles and solar PV.

can arise due to incorrect network specification. For example, the geographically defined small world networks that can be appropriate for residential solar PV may not be appropriate for electric vehicles, which are non-stationary and have different associated information channels.

Further empirical grounding of networks can be established by leveraging survey data with GIS. The responses of the surveyed individuals can be compared to the same measure in the GIS: for example, the number of solar PV systems in the neighborhood, shown in Figure 2.5. The simulation can then incorporate both measures by defining the degree distribution from the survey and the specific nodes from the GIS. Using the survey data alone would not allow for empirically defined edges to be created between nodes (connections would be the right number but randomly defined), while a static geographical model would over-estimate in-degree (too many neighbors).

2.5 Empirical Environment

Successful ABMs incorporate micro-data into the specification of the environment as well as the agents. Models that utilize GIS have a strong advantage here, as agent interactions with the environment can be modeled realistically at a high level of granularity and realism. GIS incorporation means that the model uses a Cartesian coordinate system for the study area, with datum and projection parameters maintained in the model and object attributes recorded through remote or on-site measurement. For example, the environment plays

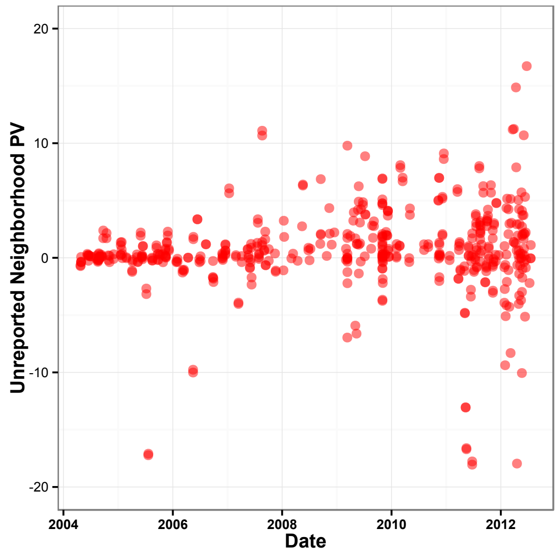


Figure 2.5: Distribution of the difference (perceived - actual) in number of homeowners in the neighborhood with PV.

an influential, though not deterministic, role in the siting of wind turbines and solar PV systems, as shown in Figure 2.6. Two examples of relevant environmental variables are shown: a 1-meter LIDAR tree-cover layer on the left panel, and transmission lines and substations right panel. It is important to note that though the environment is not autonomous, it can be highly dynamic. For example, models that incorporate land use/land cover (LULC) change allow the environment to react to changes in agent state, such as location [97]. Further, the resolution of environmental variables has been shown to have a large impact on agent behavior and ultimate conclusions drawn from the model [42].

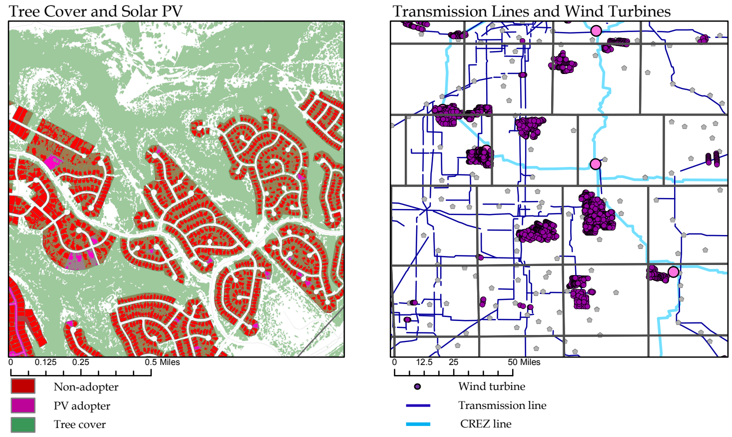


Figure 2.6: Environment variables in solar PV and wind turbine ABMs.

2.6 Precise Initialization

2.6.1 Time

The aspect of time in ABMs is critical, and has not been adequately addressed in the literature. Initializing the model in time refers to setting t_0 in the model to a specific date in an observed time series. If agent states do not reflect the ground-truth at that time initialization bias is created. Because the agent states at time t and the states at time $t+1$ are not independent, initialization bias can be perpetuated throughout the simulation.

What this means in practice is that on initialization, or t_0, the modeler should take care to set up agent and environment states using the empirical time-series. For example, if a researcher seeks to simulate the diffusion of electric vehicles in consumer markets, and has access to empirical data from 2000 - 2013, starting in 2004 might make sense. At this point there has been enough preceding activity to generate observable patterns on which to ground the model, and enough subsequent activity on which to fit and validate the model. The would then refer to matching the relevant conditions in 2004 within the modeling framework as closely as possible.

This process can be very difficult, especially if spatial validation is sought because initialized variables will have to reflect empirical spatial characteristics. However, if the model is not referenced to a specific real-world time at $t0$, fitting and validation may not be possible, and there is a good chance of simulation time that is not representative of empirical agent behavior. Improper initialization is usually revealed during validation. For example, a spike in simulated marginal adoption in the first model cycle would suggest that values were initialized too close to proposed thresholds.

2.6.2 Randomness, Independence

It is common practice for modelers to initialize assignment of agent attributes and agent location randomly (for example see [38]) regardless of how this will affect conclusions drawn from the batch. Random initialization of input variables that have an impact on output variables, or have important interaction effects will affect outcomes of the model and conclusions drawn from those outcomes. For a example of this, imagine trying to evaluate the impact of increasing electricity prices on solar PV diffusion. If attitude regarding PV and wealth are randomly and independently distributed at time 0, the impact of electricity price increase may be biased downward as there will be fewer high-income, high-attitude agents. This problem is further compounded when network effects are taken into account, given their interaction with agent attributes and spatial location. With the use of stochastic variables, sensitivity testing and the use of batch runs (see section 2.8) is critical. New techniques such as Bayesian Analysis of Computer Code Outputs (BACCO) have improved the ability of modelers to use multi-parameter models efficiently [111].

Random initialization can be used for variables for which data cannot be associated with specific agents, for example with anonymous survey data. In this case, the distribution of the variable should be identified and its parameters specified through statistical modeling. This methodology is available in a number of common statistical packages, such as R's distFit or Matlab's fitdist tools. Perhaps the best random initialization occurs through Bayesian methods, where the prior is defined by existing theory, and empirical data is represented in the likelihood function. Initialization via sampling from the posterior distribution can yield robust results that mitigate the potential for over-fitting without requiring agent-level data for $t0$. However they are created, random initialization distributions must be correctly specified and empirically grounded. For variables for which no empirical data is available, sensitivity testing on the specified distribution can provide a range of potential outcomes.

2.6.3 Input vs Emergent Variables

There is an important distinction between those variables that are important to initialize in order for the model mechanism to be specified correctly, and those measures that should emerge from a correctly specified model. In general, emergent patterns should not be initialized–built into the model directly–because it can cause over-fitting, and thus poor predictive performance. The best models will create patterns based on the behavioral rules, network configurations and resolutions, agent types, and decision thresholds. However, if not enough data is available, this fitting technique may be justified. One example of this is using regression models to estimate agent behavior, and building in probabilities to the ABM, shown in Figure 2.7. The odds of adopting solar PV increase with income, vary with the density of systems in the neighborhood. These odds could be built into the model to increase validation results. However, building the odds in directly fits the model to observed behavior rather than simulating it. This pattern should emerge from the model mechanics, such as the interaction between attitude, income, and the number of neighborhood systems.

An example of a valid input variable initialization is that of consumer attitudes. Figure 2.8 shows attitude as a multi-dimensional measure of opinion of a technology observed in empirical data across five income quintiles (Q1-Q5). The distribution of this metric can be mapped to income, allowing the measure to be empirically initialized at a the start of the model (t_0) within the ABM without over-fitting the model.

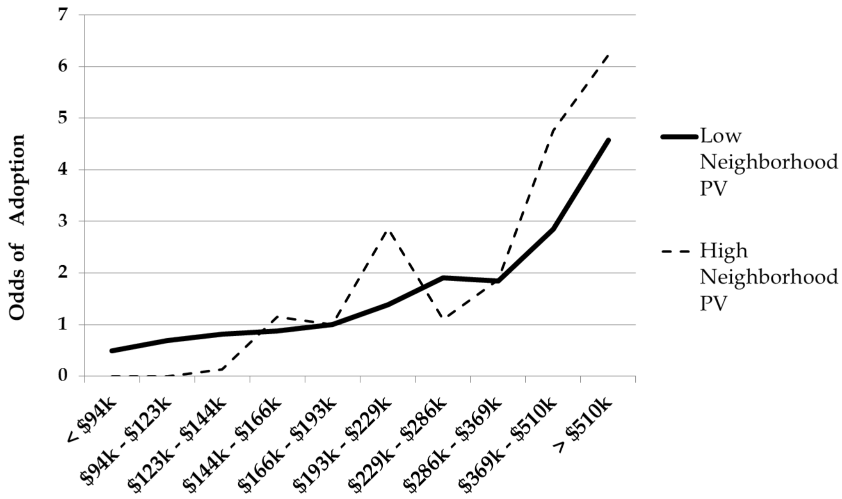


Figure 2.7: Effect of income on odds of adoption for neighborhoods with high and low numbers of PV systems.

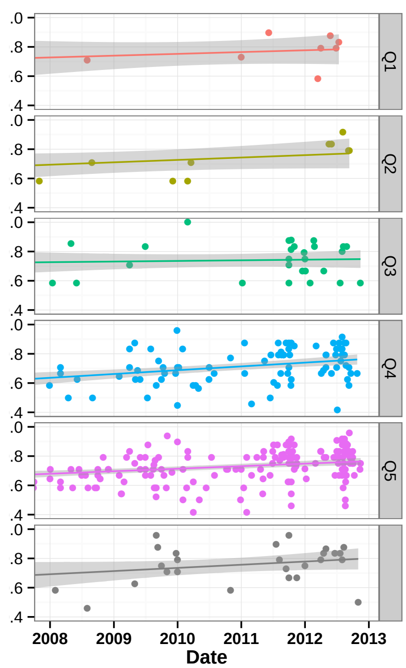


Figure 2.8: Empirical attitude evolution over time for solar PV consumers for different wealth levels.

2.7 Empirical Validation

The comparison between empirical data and simulated data should be performed on the cumulative as well as marginal distributions. The cumulative distribution will show the overall fit of the model, while the marginal distribution will reveal specific times where the model has deviated, as well as give a sense for how well the model matches the any empirical volatility. One common deviation can be due to seasonality–if there are systematic deviations on the margin, this suggests the existence of a pattern in the empirical data that has not been incorporated into the agent decision rules.

Over time, the most simple validation metric is the deviation, or root mean squared error (RMSE) between the empirical observations at time t, and the model outcome at time t. Residuals are calculated for each run within a batch along one agent-state dimension. The residuals are squared and averaged. The residuals should be randomly distributed with the central tendency around zero such that they do not show any time-trends.

Spatial validation means comparing simulation output maps against empirical maps. While most geographical comparison still relies on cell-by-cell evaluation [151] where the two maps are overlain and the error is calculated arithmetically, this method is deeply flawed for many applications [114, 148]. Two potential methods are suggested here: fuzzy numerical analysis and spatial clustering.

The fuzzy numerical statistic, Ξ is calculated in equation 4.9, where A is a simulated raster and B is the empirical raster (either density or clustering), and $w(d)$ is the function defining the weighted distance between the two cells. [59, 60].

$$\xi_i(A,B) = max^N{}_j(f(A_i,B_j)w(d_{i,j})) \tag{2.1a}$$

$$\Xi_i(A,B) = min(\xi_i(A,B,),\xi_i(B,A)) \tag{2.1b}$$

$$\Xi(A,B) = \frac{1}{n}\sum_{i=1}^{n}\Xi_i(A,B) \tag{2.1c}$$

$$f(a,b) = 1 - \frac{|a-b|}{max(|a|,|b|)} \tag{2.1d}$$

Another useful measure for spatial validation is the use of a clustering metric. Evaluation of clustering of observations (Adoption of solar PV, Electric vehicles, or the construction of wind turbines) differs from density because it takes into account the likelihood of each value given its neighboring values. Moran's I can be used to test for the presence of global clustering [77], and the Getis Ord $Gi*$ test statistic [49], and derived P values can be plotted and compared in contingency tables. If there are systematic differences in the empirical and simulated results, but good temporal fit, it is a sign of either incomplete geographic information in the model environment, poor characterization of agent interaction, or misrepresentation of networks. Figure 2.9 shows empirical and simulated data for wind turbine placement compared through density, and spatial autocorrelation metrics. The spatial over-dispersion in the simulation reflects the fact that land value is incomplete in the GIS, which allows agents to economically site turbines further apart.

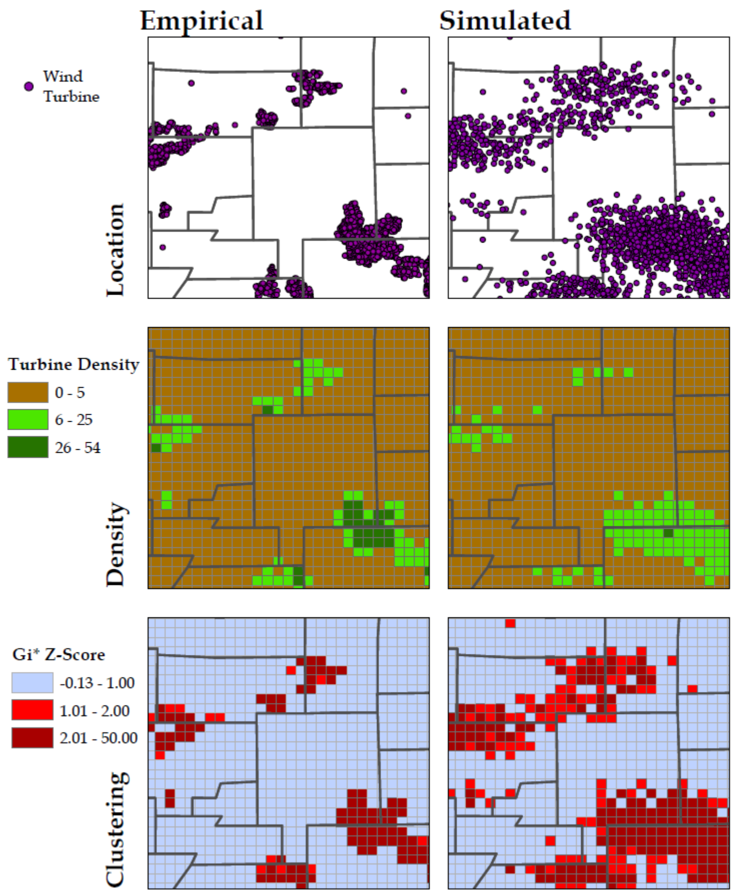


Figure 2.9: Empirical wind turbine location compared to simulated location.

2.8 Simulation Process

2.8.1 GIS Integration

The integration of Geographic Information Systems (GIS) in ABM has several advantages.

1. Most processes have a spatial component, either at the agent level or the environment level.

2. Modeling this component creates a more realistic simulation, increasing the validity of the model.

3. Modeling this component allows the model to be validated along another dimension, increasing the robustness of the model.

However, the best way to approach this integration is not always clear, after all, ABMs are explicitly dynamic and time-dependent, while robust representation of time within GIS remains a challenge [65]. Here a few key insights on the simulation process in general are offered, several software packages that integrate GIS and ABM are covered.

2.8.2 Randomization

If a system is perfectly deterministic, ABMs offer few advantages over other techniques. One of the strengths of ABM is allowing for the modeling of systems with some inherent randomness– for example, the choice of agent i to interact with agent j or agent k. The order in which the agents act (step through actions) should also be random at each time step, unless there is empirical justification for natural ordering. Critically, if agent location plays a role in agent interaction or the model dynamics, the placement of agents in the GIS should not be random, but be initialized to reflect empirical location data at $t0$. If agent interaction has a geographic component, empirical patterns should provide insight on the scope/scale of interaction. One useful technique is through Ripley's K function [37], or one of the many transformations thereof, such as the one shown in equation 2.2.

$$L(d) = \sqrt{\frac{A \sum_{i=1}^{n} \sum_{j=1}^{n} k_{i,j}}{\pi n(n-1)}}, i \neq j \tag{2.2}$$

where L is a measure of the dependency on point (incident) data as a function of distance d, the total area A, the number of data points n, and a weight k.

2.8.3 Batches

When ABMs contain random elements, runs with identical parameters (inputs) will produce different outcomes. If the ABM is run often enough, the resulting outcomes can be placed in distributions. While it is acceptable to present the central tendency and shape (deviation) of the output, when possible the full distribution should be shown in the results. Sensitivity analysis on the minimum number of runs involves holding the parameters constant, varying the number of runs, and analyzing the deviation from the outcome mean as a function of the number of runs. If this is plotted graphically with deviation on the y axis, the minimum number of runs can be chosen where the slope of the line approaches zero. Another potential difficulty that arises with Batch runs on a GIS integrated model is the spatial validation of batch data. For social-technical energy systems modeling where the outcome is a purchase decision, each run will produce agent states with binary adoption values for each time interval. Many spatial statistics procedures will require batch aggregation before the procedure is completed. In this case, the empirical state of agent i at time t: $S_{e,i,t}$ can be compared to the simulation output expected value: $E(S_{s,i,t})$.

2.8.4 Time

In an ABM, the basic unit of time is the "tick", or model cycle. Because this unit has no inherent value, this cycle must be tied to the target system units. Without this link, temporal validation is impossible. Like other aspects of the model, the time step should reflect empirical patterns observed in the target system. If there are multiple temporal patterns that are observed, this can be controlled through the inclusion of an inner loop within the larger step function. For example, agents might reassess their financial capability every month but interact once per week.

2.8.5 Number of agents

Calculating the number of agents represented in the model is simple: it should match exactly with the number of empirical individuals in the population on which the modeler has data, given the chosen level of aggregation, as discussed in section 2.2. In many cases the empirical data is a sample from the modeled population. In this case the number of agents should match the population, but have characteristics derived from the sample. If empirical data is available for a greater number of individuals than can be modeled, geographic or other empirical boundaries should be used to restrict the study area, increasing the chance of accurate representation. However, generalizations drawn from a restricted sample may not be valid.

2.8.6 Simulation package

Agent-based simulations can be run on a variety of software packages and programming languages, each with distinct advantages and disadvantages. The choice of software should reflect the project goals and the constraints of the research team. A comprehensive review of the available tools is well beyond the scope of this paper, and other reviews have covered this topic adequately [50, 121]. Instead, I note a few software packages of interest to researchers interested in explicit integration of GIS with ABM. Further review of this topic is available elsewhere [28].

1. **Agent Analyst** Utilizing the Repast Engine (see item 2), Agent Analyst interfaces directly with a ESRI's Arc GIS software and easily handles vector and raster agents as well as environments. Agent Analyst allows the use of explicit GIS processing through the ArcPy Python module developed by ESRI, and model outputs are exported directly to ARC

Info, allowing for local model outcome processing and analysis using the Arc GIS spatial statistics tools [73].

2. **RepastS** The Recursive Porous Agent Simulation Toolkit (Repast) was developed by the University of Chicago and is maintained by Argonne National Laboratory. Spatially explicit models can be built in RepastS and visualized in included 3D satellite raster display. RepastS also supports importing shapefiles for agent and environment classes.[28, 90]

3. **NetLogo** Raster and vector data are now supported in NetLogo, significantly increasing the flexibility of the software. However, substantial geoprocessing must be completed through third party applications [152].

4. **R** While R does not contain built-in ABM structure, the flexibility of the program allows for ABM functionality, and there are open-source vector and raster mapping packages available. R does not provide the GUI environment of Agent Analyst, RepastS or NetLogo, but it has the advantage of scalability, flexibility, simple parallelization, vectorization, and ease of integration in high performance computing environments.

2.9 The SECAD Model

The remainder of this thesis will be dedicated to the implementation of the Solar Energy Consumer Agent Decision (SECAD) model. The SECAD model is an empirical ABM built on the principles defined in this chapter, and seeks to simulate the diffusion of residential solar PV in Austin, Texas.

Chapter 3

Agents in the SECAD Model

3.1 Residential Households

The primary agents operating in the model are single-family residential households. The number of agents was determined by using Travis Country Appraisal District parcel data, selecting those parcels which were within the Austin Energy Service Territory, and were designated as single family residential. Under these criteria, 173,466 household agents were used in the SECAD model. Each of these households has a number of attributes, shown in table 3.1. One of the most important factors that sets the SECAD model apart from other ABMs is the degree to which these attributes are empirically derived. Because the ability to explicitly incorporate heterogeneity across multiple dimensions is a defining feature of ABM, the degree to which this heterogeneity is grounded in actual data has a large bearing on verification and validation of the model (see chapter 2).

3.2 Electric Utility

Properly, the electric utility entity is not an agent in that it is not heterogeneous–effectively it is an agent class with $N = 1$. However, it is discussed here to

Table 3.1: Heterogenous variables and attributes in the agent class.

Agent-level Attributes

Variable Name	Description	Basis
sia_i	Socially informed attitude: overall opinion of solar	Derived
pbc_i	Perceived behavioral control: economic capability	Derived
kW_i	PV system size in kW AC	Derived
s_i	Size of the home (sq. ft.)	City of Austin
${s^P}_i$	Size of the lot (sq. ft.)	TCAD
T_i	Tree cover (sq. ft.)	City of Austin LIDAR
$w1_i - w4_i$	Importance of financial, social, & environmental factors on installation decision	UT Austin Solar Survey
Ec_i	Environmental concern	UT Austin Solar Survey
$PayE_i$	Willingness to pay to protect the environment	UT Austin Solar Survey
$NeiE_i$	Level of environmental concern in the neighborhood	UT Austin Solar Survey
E_i	Index of environmental components	Derived: $f(Ec_i, PayE_i, NeiE_i)$
Pr_i	Characterization of the profitability of solar	UT Austin Solar Survey
Ms_i	Net monthly savings	UT Austin Solar Survey
PP_i	Simple payback	UT Austin Solar Survey
F_i	Index of financial components	UT Austin Solar Survey
Ac_i	Contacts with PV systems outside the neighborhood	UT Austin Solar Survey
Ns_i	Contacts with PV systems within the neighborhood	UT Austin Solar Survey
Mo_i	Degree of motivation from neighborhood systems	UT Austin Solar Survey
Con_i	Confidence from neighborhood systems	UT Austin Solar Survey
A_i	Index of non-neighborhood social components	Derived: $f(Ac_i)$
Nei_i	Index of neighborhood social components	Derived: $f(N_i, Mo_i, Con_i)$
S_i	Index of social components	Derived: $f(Nei_i, A_i)$

demonstrate the possibility of expansion of the model to multiple service areas. The utility creates an annual budget for its solar program, and gives out a rebate incentive to solar adopters in proportion to their system size: the rebate amount is described in dollars per watt, with system sizes capped at a certain level. Thus, the amount that the utility will draw from its budget depends on the adopter system size and the number of adopters. In the budget constrained case, the utility will tolerate one quarter of deficit only before shutting down the rebate program. In the unconstrained case, the utility will allow a constant deficit roll-over.

When a household is ready to adopt, the agent chooses a system size. The agent will choose a size in proportion to need.

$$kW_i = \beta_0 + \beta_1 \frac{s_i}{1000} + \beta_2 t + \beta_3 \frac{T_i}{1000} + \epsilon_i \tag{3.1}$$

After an agent is assigned a system size and price, the rebate is deducted from the utility budget for the current fiscal year (system sizes above 20kW are not eligible for a rebate) at the end of the quarter. Budget data was gathered from historical Austin Energy quarterly reports, and reflects the empirical residential solar PV budget available for each fiscal year in the simulation. Critically, in the constrained case, if the utility agent is currently in deficit, no rebate will be allocated and the household will delay the purchase decision.

Chapter 4

Methods

Agent Analyst [73], an extension of the ESRI ArcGIS® software suite utilizing the Repast engine [27], was used for verification and validation of the model on one zip-code[126]. The platform was extended in the R programming language to all residential households in Austin, TX, and simulation was performed on the Texas Advanced Computing Center's Stampede Supercomputer. In this section, the structure of the simulation process as well as the three central components of the SECAD model are described.

4.1 Model Structure and Process

The in-simulation agent decision behavior is shown diagrammatically in figure 4.2. Importantly, there are three major components to the model: the Attitudinal component, the Social component, and the Economic component. The basic relationship between the components is the following: Agents modify their attitude(sia) about solar through interactions within their social circle. If attitude is sufficiently high, agents compare the simple payback at that time period to their perceived behavioral control (pbc)– their wealth plus any physical constraints, such as tree cover over their roof. Perceived behavioral control

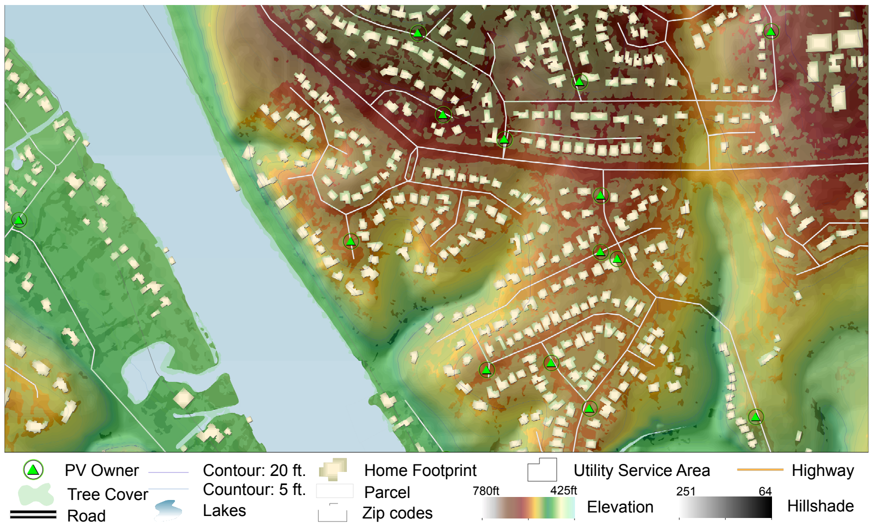


Figure 4.1: The geographic information system tied to the SECAD model.

is static, but is compared with the payback period, which is dynamic. The model environment is a detailed and multi-layered GIS, as shown in figure 4.1

Simulation batches of 100 runs with identical parameters were initialized to Q4 2007 conditions, described in section 5 and cycled forward to the 2nd quarter 2013 and validated against empirical results. The results of the validation are shown in Section 5. Goodness of fit is evaluated over space and time. Temporal validation is based on RMSE of the number of adopters predicted by the model at each time step. Spatial validation is based on simple error calculation (empirical - simulated), Fuzzy Numerical similarity statistics, and (Harr) wavelet transformation correlation coefficients–all of which evaluate the similarity between two raster maps. This raster map is derived from kernel density raster maps of solar PV installations over the study area (simulated and empirical).

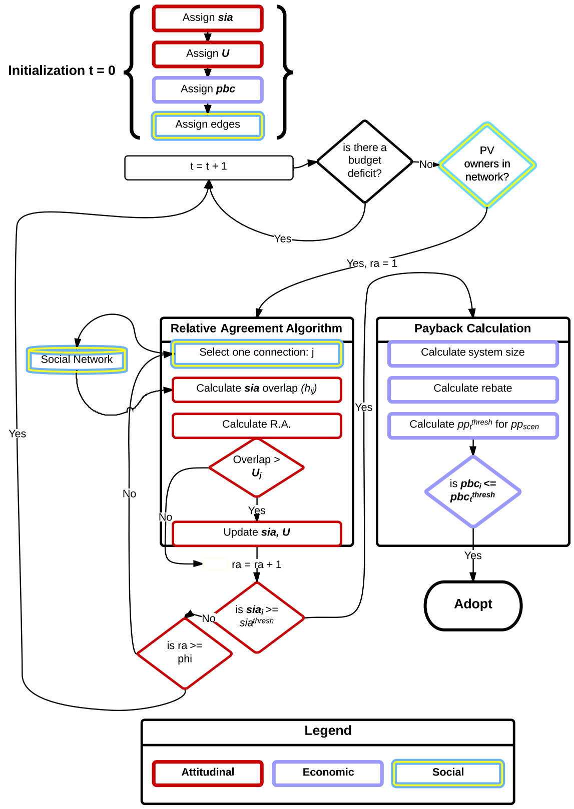


Figure 4.2: Flowchart of the behavioral rules applied to the household agent class in each time step.

Table 4.1: Agent-level variables active during the model cycle.

Agent Variables			
Variable	**Description**	**Basis**	**Dynamic**
sia	Socially informed (overall) attitude of solar PV	Empirical: see section 5.1	Yes
U	Uncertainty: confidence regarding sia	Derived: $f(U^{dist})$	Yes
pbc	Perceived behavioral control: Ability to follow intention	Empirical: $\sum_K^{k=1} (log(Z_k + 1))/K$	No
A	Solar PV adoption status	Empirical: Q4 2007	Yes

4.2 Attitudinal

To operationalize the attitude and social norms components of TPB the relative agreement algorithm is used [34, 99], an extension of bounded confidence models [63]. This algorithm is a significant improvement from other common behavioral models, the most common being probabilistic [13], number -of-neighbors [35], and percentage-of-neighbors [16]). These agent-based models do not take advantage of heterogeneity in the agent state as a compliment to heterogeneity in network position. The relative agreement algorithm is particularly well adapted for ABM because it allows for heterogeneity in state, and can create spillover and multiplier effects without the problems associated with linear models such as identification, endogenous group formation, correlative unobservables, and simultaneity [62].

Social influence in the network is modeled through two dynamic states: socially informed attitude (sia) and uncertainty (U), both of which are heterogeneous in the agent class. Socially informed attitude is initialized to reflect attitudes in the study area at t_0. Because it is next to impossible to obtain accurate data on past opinion through survey data, instead this state is modeled, as described in section 5.1. Uncertainty is distributed in proportion to the inverse absolute value of sia, as in the absence of empirical distributions, behavioral research indicates that people are more likely to process relevant information if they do not hold extreme attitudes.[26, 95]. This means that agents with very high and very low attitudes will be highly influential within their social circles. Interestingly, of the other simulations reviewed using relative agreement, none explored empirically grounded uncertainty [30, 32, 33, 44, 76, 99].

In the relative agreement algorithm, pairs of agents interact (i and j, where i influences j). Each agent is associated with two variables, opinion x_{i} (equivalent to attitude in TPB), and uncertainty u_{i}. Each agent i interacts with one other random agent j by determining the opinion overlap h_{ij}. Thus, agents are only influenced by relatively similar opinions:

$$h_{ij} = min((x_i + u_i), (x_j + u_j)) - max((x_i - u_i), (x_j - u_j)) \qquad (4.1)$$

Equation 4.1 returns the overlap of the two agent's opinion levels–the distance between their opinions. Overlap does not take into account the non-overlap–where one agent's opinion is outside the range of the other, in turn decreasing their potential for exchange. The non-overlap is subtracted from the overlap, thus yielding total agreement:

$$h_{ij} - (2u_i - h_{ij}) = 2(h_{ij} - u_i) \qquad (4.2)$$

Because uncertainty is relevant both above and below the agent's opinion level, an agent j is willing to consider an opinion from agent i in the range $2u_{\mathrm{i}}$. The relative agreement is thus:

$$\frac{2(h_{ij} - u_i)}{2u_i} = \frac{h_{ij}}{u_i} - 1 \qquad (4.3)$$

If overlap (h_{ij}) is greater than the influencing agent's uncertainty (u_{i}) then the opinion of agent j is increased or decreased by the amount of relative agreement, where μ is a constant controlling speed of convergence and weight of interaction:

$$x_j = x_j + \mu((\frac{h_{ij}}{u_i}) - 1)(x_i - x_j) \tag{4.4}$$

and uncertainty is updated:

$$u_j = u_j + \mu((\frac{h_{ij}}{u_i}) - 1)(u_i - u_j) \tag{4.5}$$

Thus, certain agents with lower uncertainty will have a stronger influence on other agent's opinion, allowing for the model to represent influential agents with both high and low opinion about solar PV [31]. This presumably adds psychological realism as individuals that are firm in their convictions will probably be more convincing to others [33, 99]. Clearly, when the agent states such as socially informed attitude (sia) and uncertainty are heterogeneous, the likelihood that agent j will be influenced by agent i depends on the choice of agent i.

If the agent is influenced positively, there is a chance that the attitude has become high enough for the agent to adopt, given the ability to do so financially. The agent's attitude is evaluated against a set threshold value after each time step, making attitude a necessary but not sufficient condition for adoption. This is a shared component with the threshold models common in social network analysis [145].

4.3 Social

Agent interactions occur through their social connections, represented by a household-level modified static geographical model rewired into a small world network. In this section, descriptive measures are presented for the network in order to provide a better understanding of how the social network structure contributes to in-simulation agent interactions. Recall that locals are derived according to empirical geographic data using a GIS of residential households in Austin, TX, where each node (household) connects to its neighbors, which are defined as those within a given radius, ρ. ρ was determined empirically using Ripley's K function[37], shown in equation 2.2. These connections are further filtered for wealth similarity, as discussed in chapter 5.

Table 4.2 compares several centrality measures for the base-case social network described here and an equivalent Erdos-Renyi random graph. For all centralization measures except for betweenness, the empirical graph shows higher centralization than the Erdos-Renyi random graph, but these differences are not large.

The betweenness distribution is shown in figure 4.3. Betweenness is the ability of a node to act as a "broker" between other nodes, such that it is located on the shortest path between those nodes. The distribution of betweenness in the base-case SECAD model is more skewed than the degree distribution.

Distribution of closeness for each agent-node is displayed in figure 4.4. This is a measurement of the normalized number of steps needed to move to every

Table 4.2: Comparison of centrality measures for empirical social networks and random networks.

Centrality Comparison		
Measure	**Empirical**	**Erdos Renyi**
In-degree Centralization	0.00034	0.00034
Eigenvector Centralization	0.99921	0.81061
Closeness Centralization	3.15758	3.15965
Betweenness Centralization	0.00002	0.00001

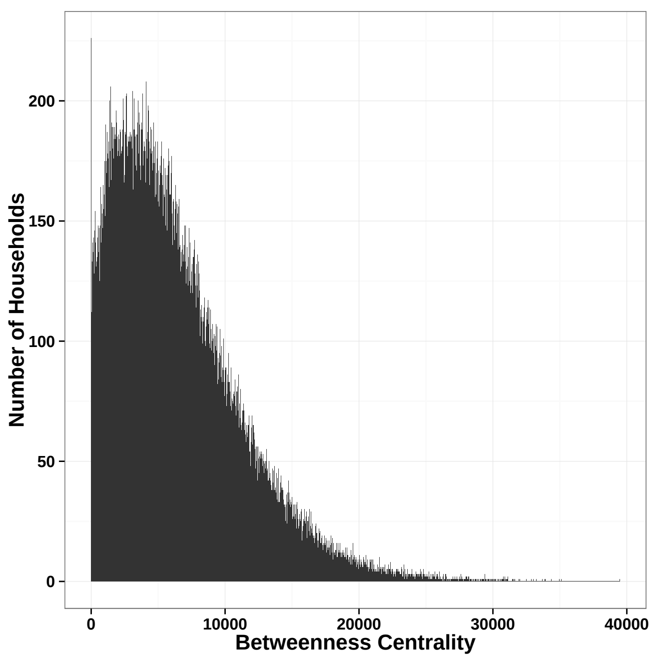


Figure 4.3: Distribution of betweenness for agents under base-case parameters.

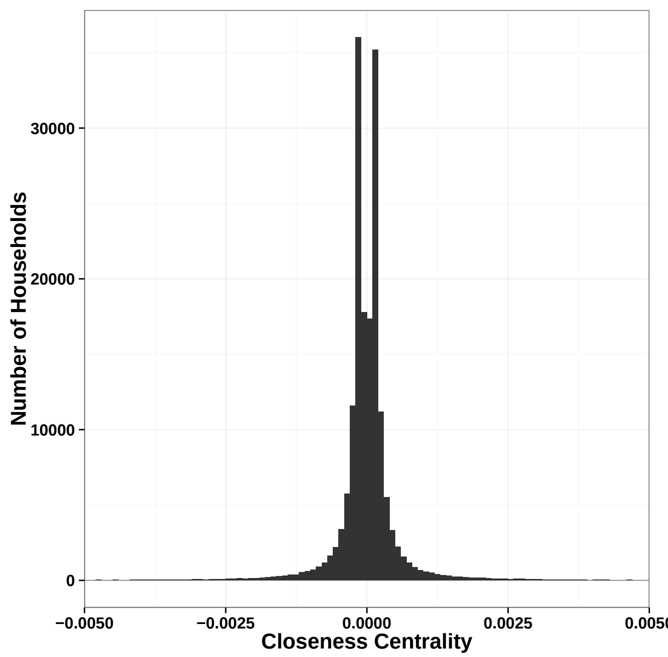
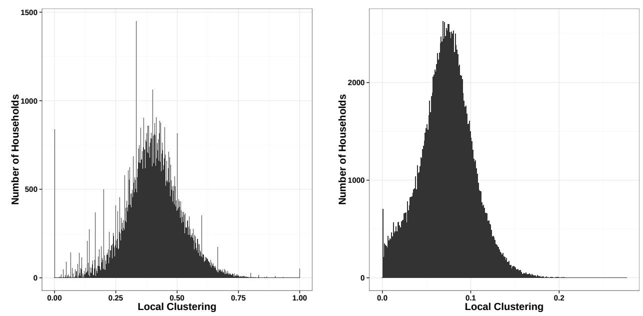


Figure 4.4: Distribution of betweenness for agents under base-case parameters.

other node on the graph from each given node. Nodes that are peripheral in the network will have low closeness.

The social network model used in the base-case has one giant component that contains all nodes. This means that all neighborhoods are connected through at least one edge, and information has the potential to diffuse from one node to the rest of the network.

The presence of local clustering is a defining property of a small-world network. Recall that a static geographical model is one where all neighbors are connected. In this model, local clustering–the probability that adjacent nodes are connected– is one. As figure 4.5 shows, local clustering is much higher



(a) SECAD Base-Case (b) Erdos-Renyi

Figure 4.5: Distribution of local clustering or transitivity for agents under base-case parameters (a), and for the equivalent Erdos-Renyi graph.

and more evenly distributed in the base-case than for the equivalent Erdos-Renyi graph. This is supported further by the global clustering coefficient (equation 1.11, calculated as 0.418 for the base-case model, and 0.083 for the Erdos-Renyi graph.

4.4 Economic

Given an attitude (*sia*) value above the threshold, agents will compare their perceived behavior control (*pbc*) to the payback period at the current time, t. Perceived behavioral control is defined as the agent's perception of their ability to follow through with an intention. For solar PV this means the agent's ability to afford the technology, subject to any limiting physical constraints (see section 5.2). Because *pbc* is taken as one index value, the assumption is that given a sufficiently financially attractive payback period, the agent will take measures to overcome physical constraints. This is empirically the case, and can be seen for example in wealthy households with high tree-cover. Presumably this is due to willingness to pay for tree-trimming services or the like.

While simple payback is a flawed financial metric in that it does not discount future cash-flows, it is the one most commonly used by potential solar adopters [119, 120]. Payback is modeled as a function of electricity prices, e_t–the value of electricity produced by the system, the price of the system in dollars per Watt DC (p), and any rebates available through utility (R) and the federal investment tax credit (ITC). Two constants, ψ and γ represent the utility discounting of the system price for rebate purposes, and the estimate of system generation in kWh respectively. The values used for these constants are the same as those used by Austin Energy in their own calculations (.9556, 1361).

$$PP_{i(t)} = (p_t - (R_t\psi) - ((p_t - (R_t\psi))ITC_t))\frac{1000}{\gamma e_t} \tag{4.6}$$

Empirically system sizes are increasing over time in the study area, while system prices (in dollars per Watt) are decreasing, as shown in figure 4.6. The agent will choose a system size in proportion to need: the size of a system (kW) is a function of the size of the home footprint (yielding available roof-space and approximating consumption), the amount of tree-cover for the roof (a constraint on available roof-space) and time in equation 4.7. On average, system size in kilowatts DC increases with time (number of quarters from Q4 2007), and with the size of the home, and decreases with tree cover, holding all other factors constant. This model was fit to empirical data for all PV owners in the study area (n = 2738). After three influential outliers (cook's D >0.4) were removed, the $Adj R^2$ was calculated as 0.32, with a p-value <0.001. All variables were highly significant (p <0.001), diagnostics showed that residuals were distributed normally, and VIFs were below three.

$$kW_i = \beta_0 + \beta_1 \frac{s_i}{1000} + \beta_2 t + \beta_3 \frac{T_i}{1000} + \epsilon_i \tag{4.7}$$

After an agent is assigned a system size and price, the rebate is deducted from the utility budget for the current fiscal year (system sizes above 20kW are not eligible for a rebate). Empirical budget data was used from historical Austin Energy quarterly reports. Critically, in the budget constrained case, if the utility agent does not have sufficient funds, no rebate will be allocated. If there is no rebate available, the household will delay the purchase decision, subject to both attitude and control values remaining above their respective thresholds. These values are still dynamic for an agent that has delayed their decision.

Normalized system prices are not correlated with household variables, but are quite predictable as a function of time. Prices were modeled using non-parametric local polynomial regression (LOESS). A moving window comprising 33% of the data was used to create subsets of the time-series system-size data. Second-order weighted least-squared polynomial regression estimates were obtained for each subset. The pseudo R^2 from this technique was 0.72,

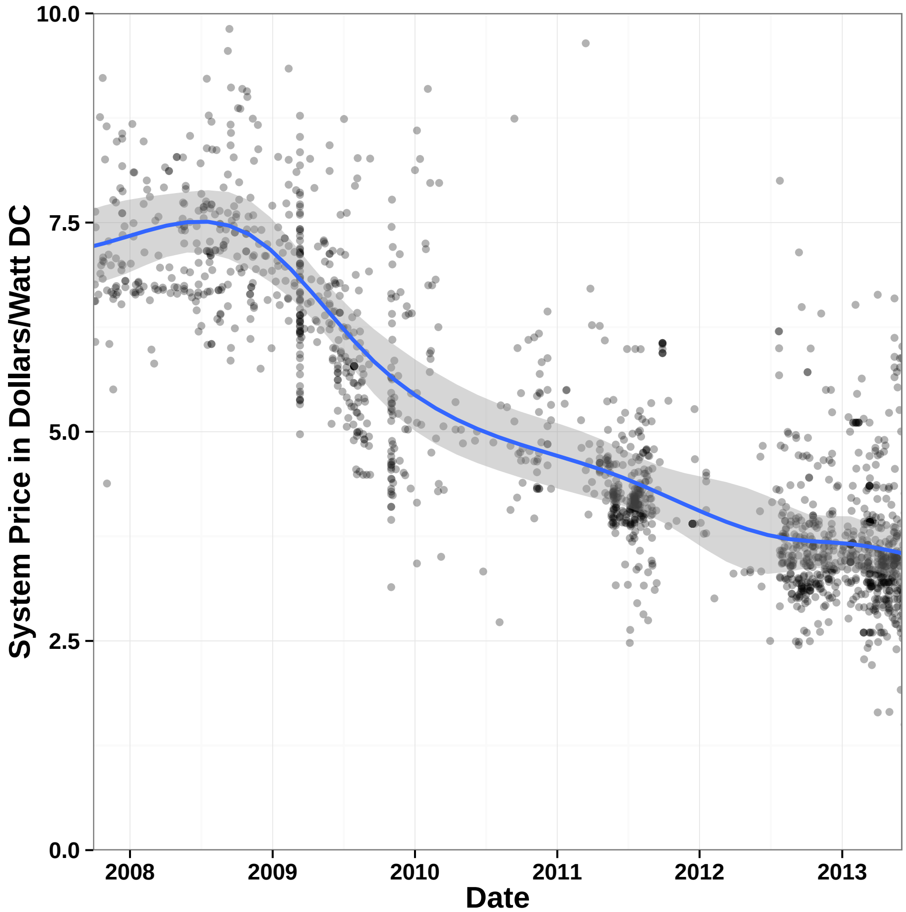


Figure 4.6: System price LOESS regression model.

and residuals were approximately normal (QQ plot showed slightly higher residuals on the right tail, demonstrating worse fit for systems that were priced much higher than the local average).

The combined economic model described in this section has the effect of limiting the number of agents that are willing to adopt during a given time period. Alone, this economic model By itself, this as if there were quasi-perfect infor-

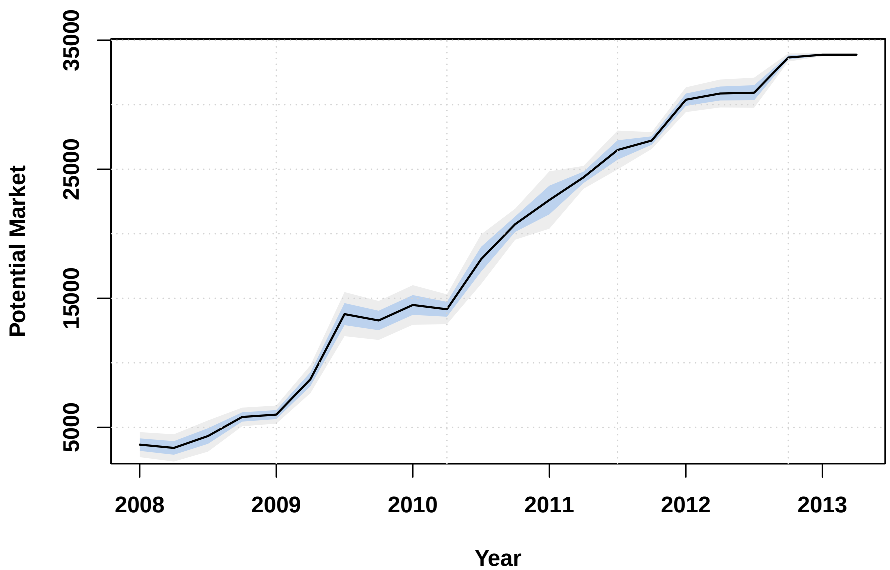


Figure 4.7: Potential market for solar PV, without informational or attitudinal constraints.

mation available to potential adopters (system prices are still variable due to the distribution of installation prices for a given time period). This is shown graphically in figure 4.7, which was created by running the economic module of the SECAD model alone. To capture the variability in system prices, the simulation was run 1000 times to generate 80% and 95% confidence intervals, shown in blue and grey.

4.5 Parameters

While additional parameters are used in the individual scenarios (described in section 7), the model has seven structural parameters used to tune the social networks, opinion convergence, and economic distribution, shown in table 4.3. Because model outcomes are sensitive to these parameters, they are used to fit the base-case model to the observed temporal data, minimizing root mean squared error. Full sensitivity testing on these parameters is reported in Appendix B.

4.6 Model Fitting and Validation

4.6.1 Temporal

Residuals for the marginal number of new adopters and the cumulative number of adopters were calculated by subtracting the model output for a given quarter from the empirical outcome in the same quarter. The deviation, or root mean squared error ($RMSE$) of the model was calculated accordingly, where q is a given quarter, $\hat{a}$ is the number of adopters in the model, and a is the number of adopters in the empirical data:

$$RMSE = \sqrt{\sum_{q=1}^{n} \left(\frac{(\hat{a} - a_q)^2}{n}\right)} \tag{4.8}$$

$RMSE$ was calculated for the marginal adoptions each quarter (the number of new adopters over that time period) and for the cumulative number of adoptions each quarter (the total number of adopters from t_0 to that period).

Table 4.3: Parameters of the SECAD model.

ABM Parameters			
Parameter	**Description**	**Base-Case Value**	**Basis**
ϕ	Number of interactions interactions per agent per tick	5	Fitted
sia^{Thresh}	Minimum *sia* value needed for agent to adopt	0.6	Fitted, relative to initialization
U^{dist}	Distribution of uncertainty in the agent population	$-1\|sia\|$	Theoretical
μ	Coefficient of convergence in RA algorithm	0.55	Fitted
ρ	Maximum distance at which an agent is considered local	2000ft	Empirical: Derived from Ripley's K function
λ^r	Percentage of connections rewired randomly	10	Fitted
$pbcC$	Maximum value used to scale *pbc* to payback	32.46	Highest observed adopter payback

The marginal $RMSE$ measures the error in volatility, while the cumulative $RMSE$ measures the overall fit. The model was fit to the empirical data during parameter sweeps.

4.6.2 Spatial

Simulation output density prediction surfaces are created by summing over the adoption outcomes for each household within each batch and dividing by the number of runs in the batch. The result is an adoption probability for each agent in the model that reflects the central tendency of the batch. A kernel density function was used to calculate the number of systems per square mile for each 100 x 100 square cell in the study area. This procedure was duplicated for the empirical data (the empirical data can be thought of as a simulation with only one run) so that the rasters could be compared directly.

The methodology for calculating error over space has received quite a bit of attention in the geography and remote sensing literature [114, 148]. While most statistical comparison of maps still relies on cell-by-cell evaluation [151], where the two maps are overlain and the error is calculated arithmetically, this method is deeply flawed for many applications. The problem with this method is that it ignores the spatial structure of errors: if the goal of a simulation is to match a pattern, for example "high, low", a structure with a similar pattern but different cell-by-cell value, for example "low, high", is actually a much better match than the error equivalent "medium, medium." Simply stated, a simulation that predicts positive for a positive target's neighbors but negative for the target is better than one that predicts negative for both, a fact not recognized by contingency-table style methods that compare cells on a direct one-to-one basis.

The fuzzy numerical statistic, Ξ is calculated in equation 4.9, where A is a simulated raster and B is the empirical raster (either density or clustering), and $w(d)$ is the function defining the weighted distance between the two cells. Here a simple exponential decay function is used over a 20 cell radius, with a half-life of five cells: $V_j = V_i 2^{-d/5}$, where d is the number of cells in the raster separating i and j. The similarity of the two values is given by $f(a,b)$ [59, 60].

$$\xi_i(A,B) = max^N{}_j(f(A_i, B_j)w(d_{i,j})) \tag{4.9a}$$

$$\Xi_i(A,B) = min(\xi_i(A,B,),\xi_i(B,A)) \tag{4.9b}$$

$$\Xi(A,B) = \frac{1}{n}\sum_{i=1}^{n}\Xi_i(A,B) \tag{4.9c}$$

$$f(a,b) = 1 - \frac{|a-b|}{max(|a|,|b|)} \tag{4.9d}$$

While the fuzzy numerical method is useful for assessing spatial similarity, the choice of parameterization of the smoothing kernel function can have a large impact on results. In order to further check the robustness of the results, we calculated correlation coefficients using wavelet verification.

Wavelet verification has gained popularity in the meteorological forecasting literature due to the need to compare forecasts against observed weather patterns [20, 24]. In this method, a wavelet transformation of the raster set is performed for several wavelets. The wavelet with the lowest Shannon Entropy is selected, and noise is removed by applying a soft threshold function, and a correlation coefficient (r) is generated. The discrete wavelets (Harr wavelets are used in this study) aggregate the rasters to coarser resolutions. Here 8th level aggregation (8x8) are used [20].

Chapter 5

Initialization

The goal of initialization is to map relevant states onto the agent class. The most common method for initialization in the literature is random initialization, where agents are assigned states from a distribution, usually uniform or normal. However, while this method is simple, it disconnects the model from the empirical reality and decreases the model's relevance. The better (but much more data and time intensive) method is to initialize the agent states to reflect the point in time at which the simulation will begin, as is discussed in Section 2.6. The SECAD model was initialized with data from TCAD, the City of Austin, and the Austin Energy Solar Rebate program from Jan. 1, 2005 to Dec. 31, 2007. At this point there were 489 solar PV adopters in the study area. These adopters were matched to corresponding survey results used in initialization.

Initialization of an agent-based model that is empirically grounded in time as well as geography presents several important challenges: specifically, the agent states must reflect the geographical and temporal ground truth a t_0 as closely as possible. Otherwise there is no reason to believe that states at $t_1 - t_T$ will be accurate; initialization bias is passed through the entire simulation, a fact which is often ignored in the literature. Ideally, empirical initialization would mean obtaining measurements of each desired state for each individual in the study area at t_0. In practice, this is often impossible. For example, to follow this methodology for our model would require surveying over 173,466 households on at least two metrics on Jan. 1, 2008.

Not having access to these empirical distributions creates two large problems in the context of ABM. First, the problem of "burn-in": we know that attitudes evolve continuously within a social context, creating networks of individuals with similar attitudes. For the model to simulate this given a random initial distribution, convergence takes a number of model steps [34]. Because the empirical amount of convergence is also unknown, the number of steps until burn-in is also unknown.

Secondly, *sia* is the "currency" that agents exchange in the model. If the distribution of this currency does not match the empirical distribution at t_0 there is a much greater chance of error over time and space. Imagine a highly connected, highly influential agent: if we measure the agent's attitude empirically, and find it to be very low, the agent's network, particularly the agent's interconnected neighbors, will also be expected to have low attitude. If this same agent in the model is initialized with high attitude, the opposite will occur in the model, creating a large residual attitude in the agent's neighborhood, and likely a high number of false-positive adoptions. To address this challenge, we instead use statistical inference, described below.

5.1 Initializing Socially Informed Attitude

Initializing attitude regarding solar PV at $t0$ (Dec. 31, 2007) poses several problems. Ideally, we would be able to survey each household in the study area, or at least a random sample of people regarding their attitude. Regardless of the usual difficulties associated with conducting a random sample survey to a large group of people, this would require the respondent to recall how they felt about a decision outcome, the choice to install solar PV or not, six years ago–which would likely result in massive measurement error. Because we did not have disaggregated information about opinion levels regarding solar among non-adopters, we instead used existing survey data to model and interpolate sia to the entire population, according to a four-step process:

i. First, we create an index that adequately describes the idea of socially informed attitude. From Ajzen [3] we use the idea of attitude as an individual's best estimation of the net outcome of an action. Over time, this attitude is modified through social interaction and expectation. Analysis of the survey data showed that attitude regarding solar was informed primarily by k primary components:

1. Financial: the profitability of the technology
2. Environmental: the environmental impact of the technology
3. Social: the perceived social norm for the technology

Thus, socially informed attitude regarding solar PV for any given individual i can be thought of as the an index of K these components ζ_k, shown in equation 5.1, where ζ_{ki} is in turn composed of financial (F_i), environmental (E_i), and social (S_i) indexes (equations 5.2). Note that this is improved by incorporating components weights in equation 5.4.

$$sia_i = \frac{1}{K}\sum_{k=1}^{K}\zeta_{ki} \tag{5.1}$$

$$F_i = PP^*{}_i + Pr_i + Ms_i \tag{5.2a}$$

$$E_i = EC_i + PayE_i + NeiE_i \tag{5.2b}$$

Relevant survey questions related to these variables in this chapter are found in Appendix A.1. $PP^*{}_i$ is the reported payback period (question 3). Pr_i is an individual's characterization of the profitability of the system (question 2). Ms_i is an individual's net monthly savings (question 3). $PP^*{}_i$ is the simple payback, estimated by the respondent as a part of their financial decision-making (question 3). EC_i is the level of overall environmental concern (question 7). $PayE_i$ is the amount an individual is willing to pay to protect the environment (question 8), and $NeiE_i$ is the level of concern for environmental issues in the individual's neighborhood (question 9).

The treatment of S_i requires some additional attention: there is heterogeneity in where potential adopters get social norms. We expect most of this to come from within neighborhoods except in areas with low system density. To allow for this, we measure social aspects from neighbors and acquaintances, and take the social norm to be the maximum of the two. Thus, δ is an indicator variable: $\delta_i = I(Nei_i > A_i)$ operating on the definition of S_i:

$$S_i = \frac{(\sum_{p=1}^{k} Nei_{pi})^{\delta}}{k} \frac{(\sum_{q=1}^{m} A_{qi})^{1-\delta}}{m} \tag{5.3a}$$

$$Nei_{pi} = (log(Ns_i + 1), Mo_i, Con_i) \tag{5.3b}$$

$$A_{qi} = log(Ac_i + 1) \tag{5.3c}$$

Ns_i is the number of reported systems in the neighborhood (question 5). Mo_i is the degree of motivation obtained from neighborhood systems (question 6. Con_i is the degree of confidence obtained from neighborhood systems (question 6). Ac_i is the number of contacts with systems outside the neighborhood (question 4). The variables k and m represent the number of components in each index p and q. Thus the definition S_i is flexible to the social characteristics observed for each individual.

ii. Equation 5.1 does not allow for observed consumer segmentation in the residential solar market, suggesting that attitudes are homogeneously constructed over the vector of individuals (apart from the social component S_i as shown in equation 5.3). In other words, ζ_k takes equal weight from financial, environmental, and social components for all agents. This assumption can be avoided by taking advantage of additional survey information. The resulting calculation is a weighted average, where the weights are the revealed importance of each component–taken from survey question 1. This is shown in expanded form in equation 5.4:

$$sia_i = w1_i F_i + w2_i E_i + (w3_i \frac{(\sum_{p=1}^{k} Nei_{pi})^{\delta}}{k})(w4_i \frac{(\sum_{q=1}^{m} A_{qi})^{1-\delta}}{m}) \tag{5.4}$$

Using the revealed weights that individuals place on different factors adds an additional layer of realism. For example, this allows for the situation where two agents can care about the environment, but it was only an important factor in the installation decision for one of them.

iii. As noted above, collecting past micro-data from the entire population of N households is not possible and would likely involve a very high degree of observation error. To infer these values we use the predictive linear regression model in equation 5.5.

$$\begin{aligned} sia_i = {} & \beta_0 + \beta_1 log({s^p}_i) + \beta_2 log({s^p}_i)^2 + \beta_3 log({s^p}_i)^3 \\ & + \beta_4(\frac{T_i}{{s^p}_i}) + \beta_5(\frac{T_i}{{s^p}_i})^2 + \beta_6(\frac{T_i}{{s^p}_i})^3 + \beta_7(\frac{{s^p}_i}{W_i})^2 + \beta_8(\frac{{s^p}_i}{W_i})^3 + \epsilon_i \end{aligned} \tag{5.5}$$

For each household i, s^p is the size of the lot in square feet, T is the amount of tree cover on the household parcel, W is the household wealth, approximated through market home value, and ϵ is the error term. The model fits well given the high variability inherent in survey data, and attitudinal measures especially. The $AdjR^2$ was calculated as 0.19. Full diagnostics and additional information about this model can be found in A.2.

iv. While the regression model presented above provides fair estimates of predicted *sia* values ($\hat{sia_i}$), it assumes that there is no geographic relationship in *sia*. This relationship is expected for two reasons: first due to potential unobserved but geographically related socio-demographic variables, and second because *sia* is *socially informed attitude*, attitudes will tend to converge locally. To account for this, we model the error term (ϵ_i) in a spatial autocorrelation model (kriging with trend). Survey data were matched to program data through a fuzzy logic procedure on names and email addresses. Addresses were then geocoded, creating precise latitude and longitude vectors for each survey respondent. These values locations associated with the corresponding ϵ and used in the kriging model.

$$\epsilon^*(u) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha(u)\epsilon(u_\alpha) \tag{5.6}$$

such that:

$$\sum_{\alpha=1}^{n(u)} \lambda_\alpha(u) = 1 \tag{5.7}$$

while minimizing L, the error variance plus an additional Lagrange parameter $mu_\epsilon(u)$:

$$L = {\sigma^2}_\epsilon(u) + 2\mu_\epsilon(u)(1 - \sum_{\alpha=1}^{n(u)} \lambda_\alpha(u)) \tag{5.8}$$

Here $\epsilon^*(u)$ is the error term from equation 5.5 at location u, defined by longitude and latitude vectors; $m(u)$ is the expected value at location u, defined by a second order spatial trend, shown in equation 5.9.

$$m(u) = m(x,y) = \beta_0 + \beta_1 x + \beta_2 x^2 + b\eta_3 y + \ beta_4 y^2 \tag{5.9}$$

$n(u)$ is the number of local (neighboring) observations at location u; α is the index of the $n(u)$ local observation locations, and *lambda* is the kriging weight, which is associated with each neighbor α at location u. λ and α are derived from the semivariogram, which models the residual value on ϵ after the removal of the local trend. A Stable model for the semivariance with six lags, a partial sill of 0.3, and no nugget was used in this model. Kriging acts as a spatial interpolator, using the sum of neighboring de-trended values, weighted by distance to location u and distance to other neighbors, to estimate values at unobserved locations [54]. One of the main advantages of kriging, especially within a simulation framework such as ABM, is that is allows for the computation not only of estimates, but of variances around those estimates:

$$\sigma^2{}_{\epsilon}(u) = C(0) - \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u) C(u_{\alpha} - u) - \mu_{\epsilon}(u) \tag{5.10}$$

The geographic distribution of standard errors from the kriging estimate are shown in figure 5.1. This map effectively shows the error that is a function of neither geography nor the variables in equation 5.5. As can be expected, error increases around the edges of the map, and where survey samples were not abundant. The green points on the map show the locations of the pre-2008 adopter sample. Figure 5.2 displays the total error adjustment values obtained by the kriging model. As the figure demonstrates, there are pockets of high and low adjustment in the study area. For example, equation 5.5 overestimates *sia* for households in the SSE and WNW, and underestimates *sia* for households in the south central and east. These values were mapped

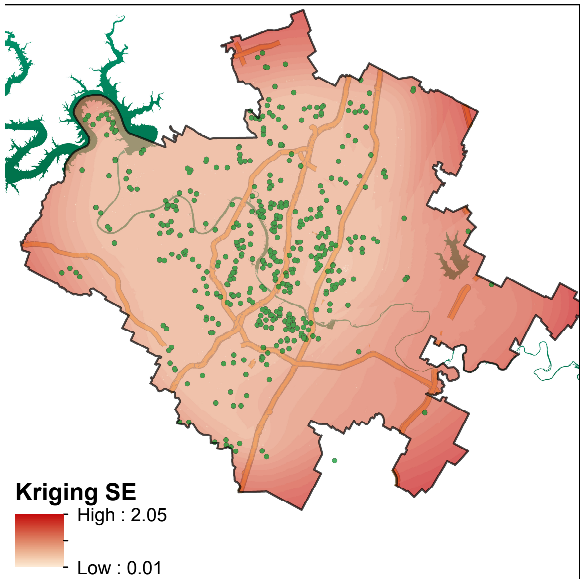


Figure 5.1: Distribution of the kriging model standard error.

onto the household agents using the latitude and longitude of the home. The individual adjustment values were then plugged in to equation 5.5, increasing the model $adjR^2$ by an additional 0.15

Thus at $t0$, sia is distributed across all agents as a function of demographic, physical, and geographic characteristics. The scaled distribution of sia is shown in figure 5.3, with the green vertical line showing the sia threshold for adoption.

5.2 Initializing Perceived Behavioral Control

Perceived behavioral control (pbc) is the degree to which an agent feels he/she is able to perform an intended behavior. Specifically, agents look to their finan-

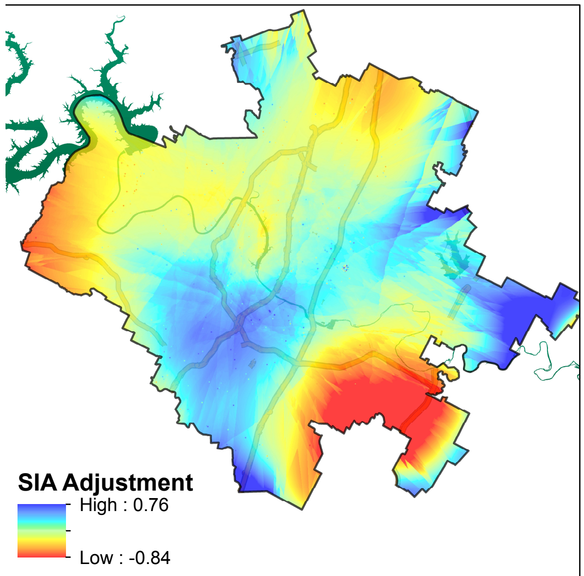


Figure 5.2: Distribution of spatial error correction values from the kriging model.

cial well-being and physical constraints and compare this with the empirical payback at the current time period. Thus the *pbc* for each agent i is:

$$pbc_i = \frac{\sum_{k=1}^{K} log(Z_k)}{K} \tag{5.11a}$$

$$s^2, W, (T-s)^2 \in Z_k \tag{5.11b}$$

Note that the components Z_k are taken on the log scale to moderate the effect of the large skew in the distribution of population variables such as home value and size. This variable was then scaled to the range of observed payback periods PP_i, allowing it to be directly compared with calculated payback for a given system price at each time t in the simulation. The choice of the scaling ceiling deserves some attention: we want to choose a ceiling that best reflects

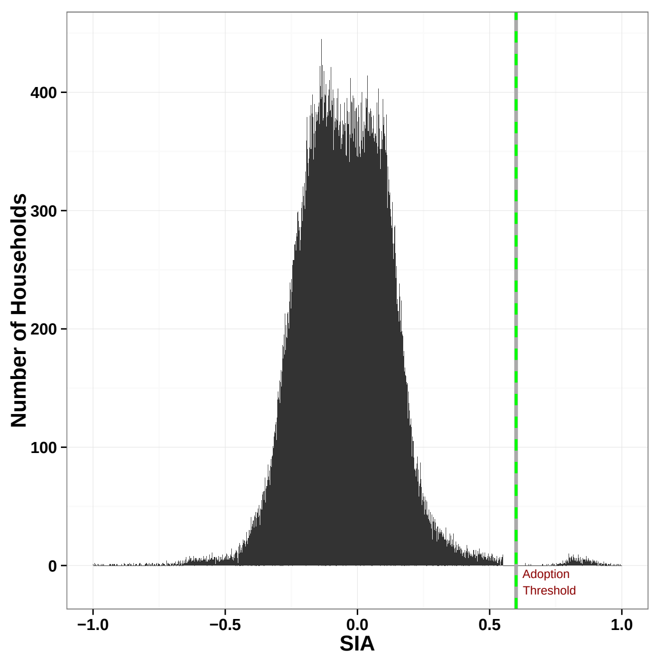


Figure 5.3: Empirical distribution of socially informed attitude regarding solar PV.

the relationship between payback and *pbc*. Here we use the ceiling of 32.46, the highest observed payback in the empirical adopter sample. In order to measure the accuracy of the scale, we can compare the empirical payback for each adopter to their scaled *pbc* value at their time of adoption. If the scale is correct, *pbc* will be greater than the observed payback at the time of adoption. As the ceiling increases, the percentage of true positives will increase, but so will the number of false positives: an infinite ceiling would pose no constraint regardless of payback. The ceiling of 32.46 results in 82.4% correct predictions, shown in figure 5.4, and results in the distribution shown in figure 5.5. Model sensitivity to this parameter is explored in section B.

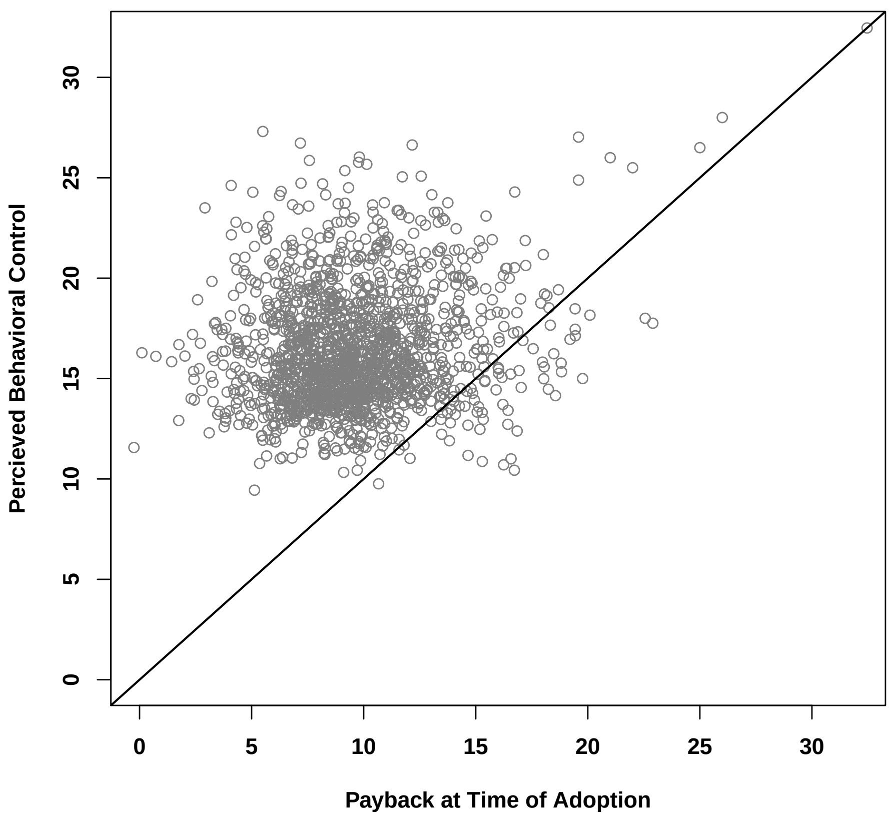


Figure 5.4: Distribution of perceived behavioral control pbc_i for each household in the simulation.

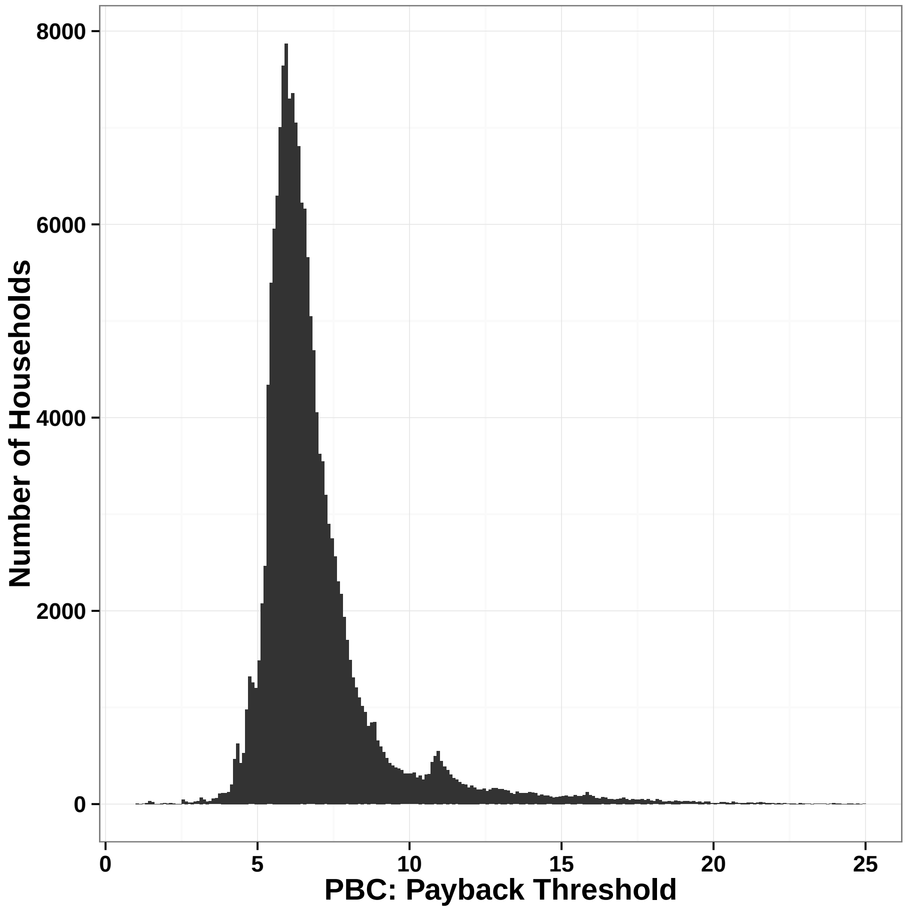


Figure 5.5: Scaled perceived behavioral control (*pbc*) for empirical adopters compared to their actual payback period.

5.3 Initializing Social Networks

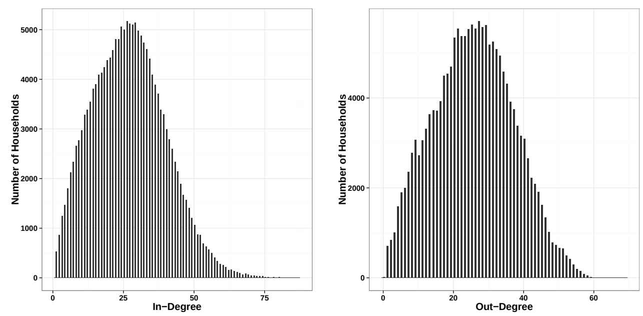
We start with the assumption that the social networks that arise from sharing information about solar ABM are entirely local geographically. This assumption is not unreasonable given that with regards to solar PV, informed neighbors are a critical source of trustworthy, relevant information. However, it is clear that neighborhoods can be defined in various ways: for example by continuity, natural boundaries, or radial distance (this has been examined experimentally in a previous iteration of this model at the zip-code level [126]). Contiguity (next-door neighbors only) is difficult to justify, while natural barriers require subjective definitions and thus reduce the potential for replication of the methodology. Radial distance is a better option but requires a radius (r) to be defined. This can be done empirically by choosing the distance d at which $L(d)$ is the greatest relative to $L(d)$ calculated for randomly dispersed points in the study area. This is shown in equation 5.12c.

$$L(d) = \sqrt{\frac{A\sum_{i=1}^{n}\sum_{j=1}^{n} k_{i,j}}{\pi n(n-1)}} \tag{5.12a}$$

$$L(d)_{rand} = \sqrt{\frac{A\sum_{irand=1}^{n_{rand}}\sum_{jrand=1}^{n_{rand}} k_{irand,jrand}}{\pi n_{rand}(n_{rand}-1)}} \tag{5.12b}$$

$$r = max(L(d) - L(d)_{rand}) \tag{5.12c}$$

Then agent i's connections are defined by those n_{local} agents within radius r. In our study area, the static geographical network model yields an approximately normal degree distribution (truncated at 0) with mean 498, and standard deviation 226.7. To scale down the degree distribution to a more realistic range, we add the additional constraint of wealth similarity. In the base-case, we maintain proportionality by calculating the squared difference in wealth between the target node and its geographical neighbors: $(W_i - W_j)^2$, and connecting those which are in the bottom 5%. Thus geographic neighbors that



(a) In-Degree (b) Out-Degree

Figure 5.6: The distribution of in-degree, the number of edges connecting to each node in a directed graph (a), and out-degree, the number of originating edges for each node in a directed graph (b).

are the most similar in wealth are connected as locals. Random connections are substituted with a 10% re-wiring probability in the base-case. This yields the degree distribution shown in figure5.6.

Chapter 6

Results

In this section results and verification are presented for base-case models over time and space. Scenario analysis results are reported in section 7, and sensitivity testing is covered in section B. It should be noted that parameters (table 4.3) were used to fit the model over time, but the spatial structure was not used in fitting.

6.1 Temporal Evaluation

The SECAD model advances ABM methodology in two important ways: first we use households-level agents at a large scale. Secondly, we have attempted to empirically initialize as many agent variables and model parameters as possible, making the SECAD model perhaps the most empirically grounded large-scale ABM attempted to date. Here we include results with random components as well as the fully empirical base-cases in order to demonstrate the effect on the model outcomes. The best fitting models were used as base-cases, shown in figure 6.1.

Because we advance the principles of empirical ABM, it is important to eval-

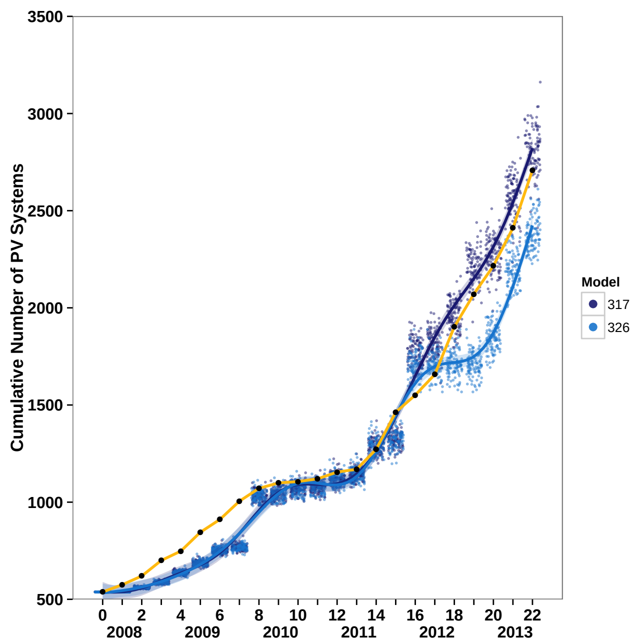
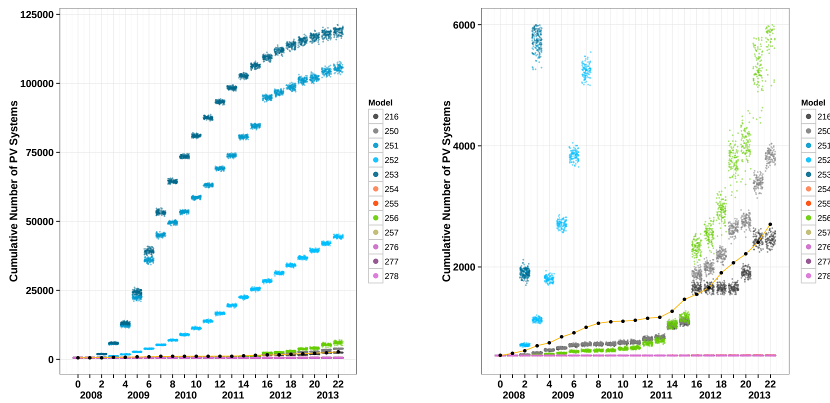


Figure 6.1: Cumulative number of installations over time, with each point as the quarterly outcome for one simulation. The dark blue represents the unconstrained budget case, and the the lighter blue represents the unconstrained case. The yellow curve and heavy black dots show empirical adoption levels.



(a) (b)

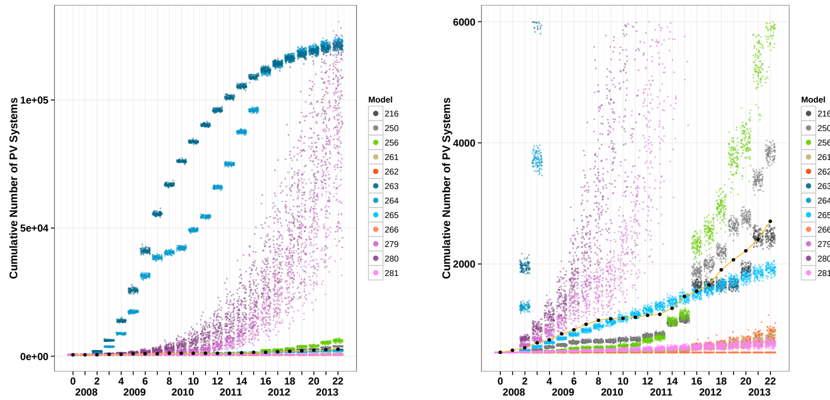
Figure 6.2: Effect on temporal model outcomes from incorporating random uniform components.

uate the performance of the empirical model against models with varying amounts of randomness. Further, this demonstrates where empirical specification is the most important. Random batches have all other parameters equal to batch 317, which is the 'unconstrained' base-case. Random components were added combinatorially, and different parameterizations of distributions were attempted. Deviation is reported in table 6.1 for random uniform distributions, and table 6.2 for unimodal distributions (Beta, Normal, and Poisson). Figures 6.2 and 6.3 show a selection of batch runs against the empirical cumulative adoption levels in the study area–showing the overall error as well as the sign and temporal location of deviation.

Through this iterative process, a fully random model was fit to the empirical data by minimizing cumulative and marginal RMSE. Batch 331 showed the best fit of any model with entirely random components tested. *sia* was

Table 6.1: RMSE calculated for marginal (quarterly) and cumulative adoption rates, base-case against random uniform initializations. Description shows random components in each batch.

	RMSE		
Batch	Marginal Adoption	Cumulative Adoption	Description
317	106.01	126.65	un-constrained
326	116.98	173.95	constrained
251	5675.18	68168.25	*pbc*
252	2080.65	21395.00	*pp*
253	6660.13	82439.46	*pbc*, *pp*
254	122.30	967.23	U
255	122.30	967.23	*sia*
276	122.29	966.60	*sia*, *pbc*
277	122.31	966.36	*sia*, *pbc*, *pp*
278	122.29	966.58	*sia*, *pp*
256	364.50	1143.23	λ^r
257	122.30	967.23	λ^r, *sia*
258	122.30	967.23	λ^r, *sia*, U
259	122.30	967.23	λ^r, *sia*, U, *pp*
260	122.30	967.23	λ^r, *sia*, U, *pp*, *pbc*



(a) (b)

Figure 6.3: Effect on temporal model outcomes from incorporating random unimodal components.

initialized to a Normal (0.2,0.3), payback was described by a Poisson (13) distribution, *pbc* was initialized to a Poisson (3) distribution, uncertainty was initialized to a Beta (5,5) distribution, and λ^r was initialized to 1 (Erdos-Renyi random graph). While this batch shows the lowest temporal error of any fully random model, it does not adequately replicate the shape of the empirical curve, most noticeably during the period 2008-2012 (see figure 6.4). Average spatial error is summarized in table 6.3 for base-case models and several important random batches (see table 6.1 and 6.2 for model descriptions).

6.2 Spatial Evaluation

After the model is fit, it is important the simulated model outcomes match patterns in the data that were not used in the fitting process. The density distribution of systems over space is one important pattern to match, given

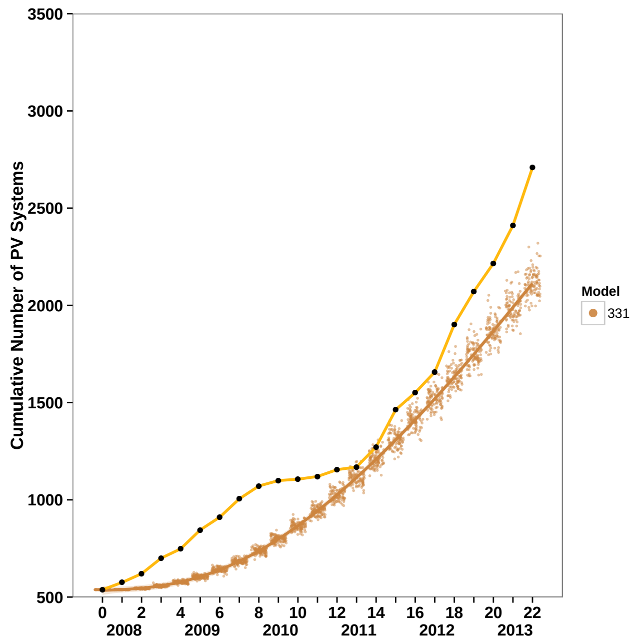


Figure 6.4: Random initialization batch with fit distribution parameters.

Table 6.2: RMSE calculated for marginal (quarterly) and cumulative adoption rates, base-case against random Normal, Poisson, and Beta initializations.

RMSE			
Batch	Marginal Adoption	Cumulative Adoption	Description
317	106.01	126.65	un-constrained
326	116.98	173.95	constrained
264	6937.44	75383.57	*pbc* Pois(13)
265	81.65	251.33	*pp* Pois(13)
263	6841.48	84472.24	*pbc* Pois(13), *pp* Pois(13)
266	122.27	966.22	U Beta(1,2)
262	106.8	874.35	*sia* N(0, 0.3)
256	364.50	1143.23	λ^r
261	109.79	885.97	λ^r, *sia* N(0,0.3)
279	6498.72	29253.84	*sia*, *pbc*
280	6190.95	37642.05	*sia*, *pbc*, *pp*
281	116.64	892.63	*sia*, *pp*
260	122.30	967.23	λ^r, *sia*, U, *pp*, *pbc*
331	65.95	257.49	λ^r, *sia*, U, *pp*, *pbc*

the importance of location for value of solar calculations, infrastructure investments and system performance among other things. Spatial error aligns with temporal error in most cases, in that lower values are found for the base-cases than for batches with random components. Relative to the fitted random model, an increase of 0.1 in the spatial correlation coefficient was observed in the base-case model. This is substantiated by the 0.08 increase in fuzzy numerical similarity. The base-case is compared to the empirical spatial distribution of PV owners in figure 6.5. Also, the wavelet analysis correlation coefficients seem to match well with the Fuzzy Numerical statistics. Both suggest that empirical distributions generate outcomes that are more similar to observed spatial patterns than random distributions, even in the fitted case. It is also notable however, that the fitted random case performs considerably better than the unfitted cases.

Table 6.3: Means of spatial error statistics for base-cases and random models.

Spatial Error			
Batch	$b_i - a_i$	Correlation, r	Fuzzy Numerical, Ξ
317	-0.99	0.56	0.42
326	0.79	0.56	0.41
262	4.20	0.39	0.34
256	-9.52	0.48	0.38
263	-158.18	0.33	0.18
280	-237.19	0.31	0.13
331	1.09	0.46	0.34

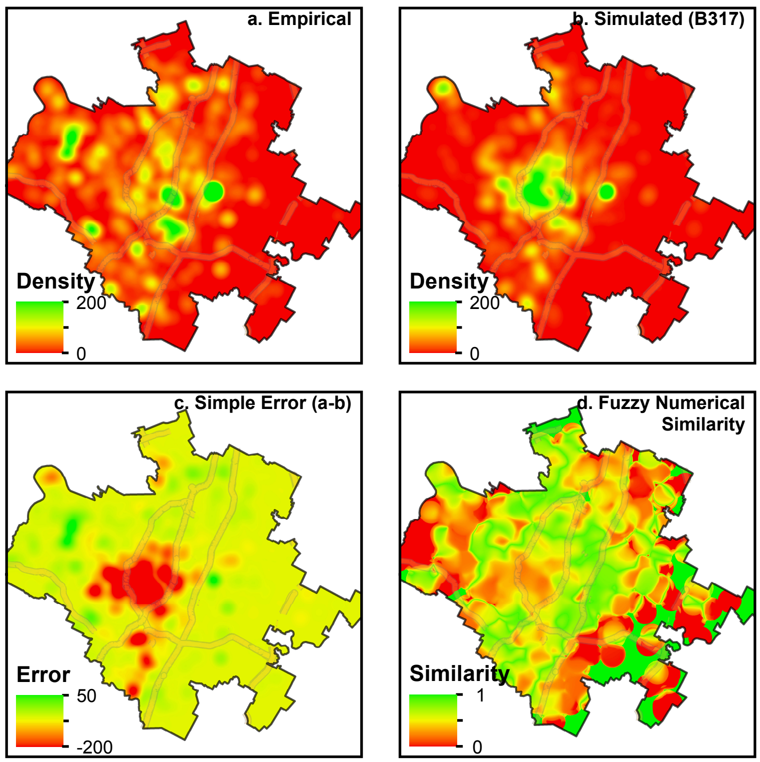


Figure 6.5: Spatial verification of the base-case SECAD model.

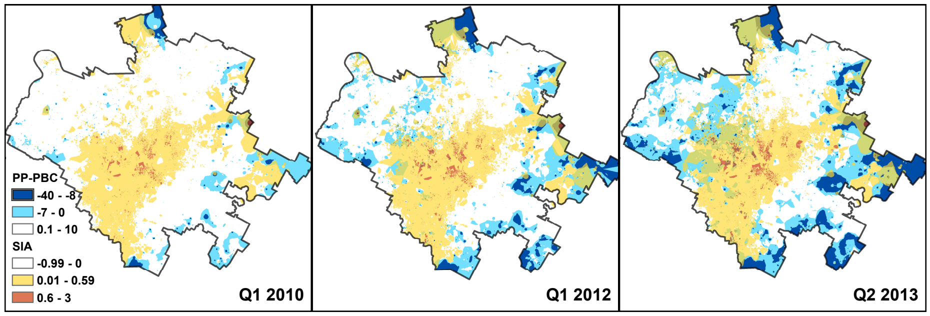


Figure 6.6: In-simulation evolution of attitude and economic capability.

6.2.1 Spatial Evolution of Agent Variables

Both *sia* and *pbc* are necessary conditions for adoption: an agent must have an attitude over 0.6, and a *pbc* that is greater than the simple payback at time t. Figure 6.6 shows the difference in payback and *pbc* in blue, and *sia* in red and yellow at snapshots taken at Q1 2010, Q1 2012 and Q2 2013 for the unconstrained base-case, batch 317. Negative numbers (blue) show areas where the agents are over the economic threshold and are thus able to adopt given a high *sia*. Positive numbers over 0.6 (red) show areas where agents are over the attitude threshold and are able to adopt given favorable economics. Overlapping areas are solar PV adopters.

Chapter 7

Scenarios

In this section we discuss several hypothetical policy scenarios, leveraging the ability of a validated ABM to simulate situations outside the observed range of the data. While we explore three scenarios here (Targeted Information Dissemination Campaigns, Tiered Rebates, and Alternative Rebate Schedules), there is large potential for expansion. The exploration of additional scenarios will be the primary focus of future work with the SECAD model.

7.1 Targeted Information Dissemination Campaigns

As shown in figure 4.7, the market penetration of PV is far beneath the potential market in terms of economic capacity to adopt; there are many more households that could afford solar PV than purchase the technology. This gap then is a function of attitude and information. This is in line with the identification of informational costs with non-technical barriers to solar energy market penetration [96]. It has been noted in the literature that information dissemination campaigns and direct subsidies act as economic complements rather than substitutes: the combined effect of information dissemination campaigns and subsidies has a greater effect than either enacted in isolation [7]. Thus it would appear that there is large potential for identifying influential nodes in the network through centrality measures, and using them to diffuse information through the social network.

Existing solar policies are targeted exclusively at the economic barriers to solar PV adoption. In this scenario, we reroute a small portion of the utility's budget into an informational campaign. These kinds of campaigns are often performed by solar community organizations (SCO's), and can include workshops, meetings, mailings, home visits, etc.[107]. While we leave the exact nature of the campaign unspecified (in reality this campaign could be any one of or any combination of the above), we assume that the effect of the campaign on the target household is a marginal decrease in uncertainty (U).

To operationalize the effect of an added informational campaign, it is necessary to estimate the cost of this intervention. There is little information about the cost-range of these types of programs. One review of the medical intervention literature found costs ranging from \$0.03-\$0.32 per person per year using mass-media resources. However, targeted information dissemination campaigns presumably require substantially greater investment. Here, we assume a highly conservative cost of \$10 per 0.01 decrease in uncertainty. No literature was found that attempted to quantify the cost and return associated with this kind of campaign in terms of attitude or uncertainty. The cost per agent, the number of agents, and the method for determining the targeted agents are all scenario parameters. Here we hold the cost and the marginal uncertainty decrease constant, and explore the number of targets and mode

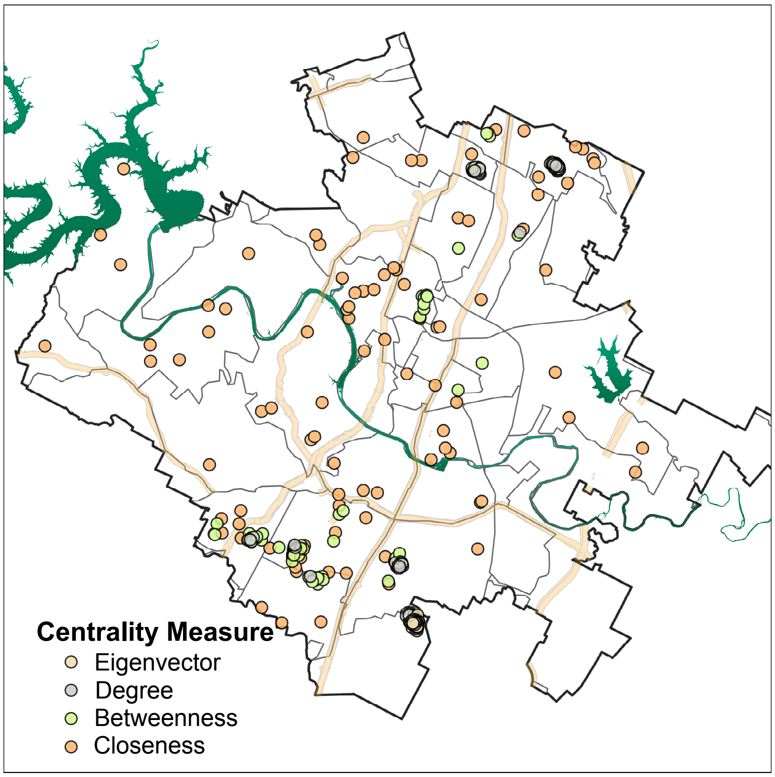


Figure 7.1: Geographic distribution of central agents in the study area according to different centrality measures.

of targeting. Specifically, we target specific households based on their centrality in the social network. Figure 7.1 plots the 100 most influential nodes in the network with respect to each of the centrality measures discussed in section 1.6. These targeting methods are compared to randomly chosen target agents (Batches 351, 352), and results are summarized in table 7.1. Percentage increase reflects the mean difference in cumulative installations relative to unconstrained base-case at end of simulation (Q2 2013).

Table 7.1: Targeted information dissemination campaigns outcome summary.

Targeted Information				
Batch	% Increase	$ per Watt	$ per Watt Increase	Description
317	0.00	2.04	0.00	Base-case, un-constrained
351	2.11	2.05	0.01	100 random targets
352	9.34	2.09	0.05	random targets
310	-0.01	2.05	0.01	100 in-degree targets
311	1.39	2.07	0.03	500 in-degree targets
312	10.34	2.10	0.06	100 betweenness targets
313	12.03	2.12	0.08	500 betweenness targets
322	1.75	2.06	0.02	100 closeness targets
323	9.54	2.09	0.05	500 closeness targets
324	0.08	2.05	0.01	100 eigenvector targets
325	3.39	2.06	0.02	500 eigenvector targets

7.1.1 Degree Centrality

We targeted 100 influential agents based on highest degree-centrality. We hypothesized that because these agents were the ones most likely to be reached out to, seeding them with information would result in an increase in cumulative

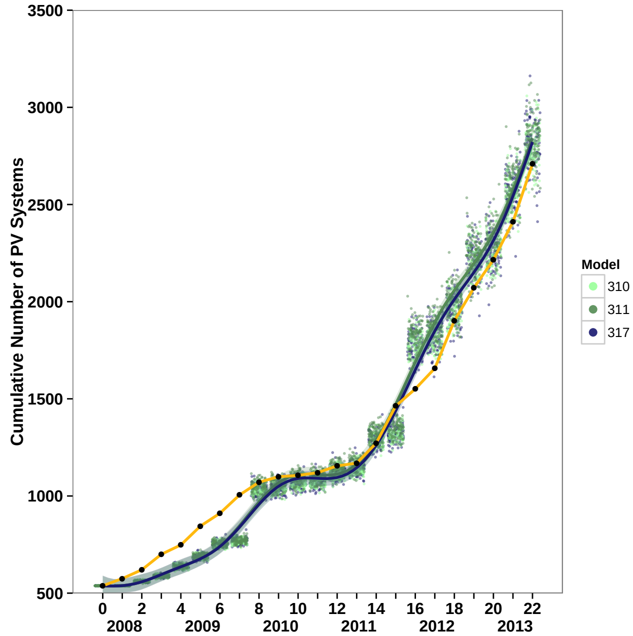


Figure 7.2: Informational dissemination campaign simulation targeting high in-degree agents.

adoptions. However, as figure 7.2 shows, there was no noticeable increase off the base-case, even when the number of targeted agents was increased to 500 (Batch 311). This is explained by the dynamic illustrated in figure 7.3–there is a clear impact on *sia*, but because central nodes tend to be co-located in highly dense neighborhoods with lower *pbc* levels, there is little impact in terms of installed systems.

7.1.2 Betweenness Centrality

Next we targeted influential nodes based on betweenness centrality in batches 312 and 313. Agents with higher betweenness tended to be less clustered

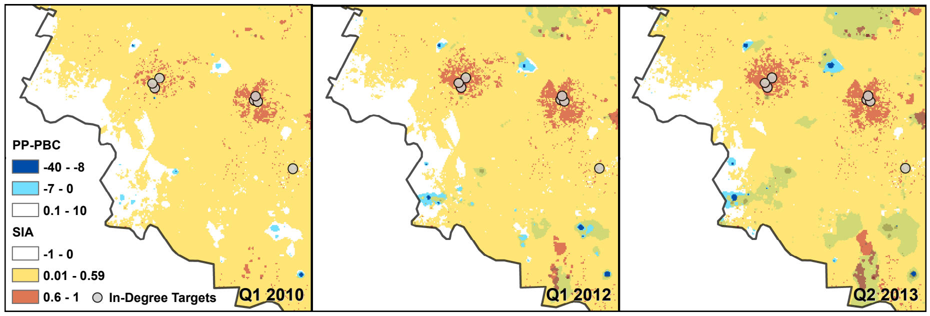


Figure 7.3: Evolution of attitudes and economic capacity under informational dissemination campaign scenario targeting high in-degree nodes.

than those with high in-degree, increasing their potential for impact. Further, these agents are by definition able to act as connectors between different neighborhoods. Here we did observe a relatively large increase in the cumulative number of installations over the base-case (see table 7.1). This is due to high-betweenness agents with connections to areas that become economically activated as solar PV prices decline (figure 7.5).

7.1.3 Closeness Centrality

The nodes with highest closeness centrality are those which can reach all the other nodes on the graph in the fewest amount of steps. They tend to be well-dispersed geographically. Batch 322 examines the effects of an informational campaign targeted at the 100 a nodes with the highest closeness centrality. We find an increase the cumulative number of installations in the study area as a result of this campaign. The effect is increased in Batch 323, which targets 500 agents based on closeness. Results are shown in figure 7.6.

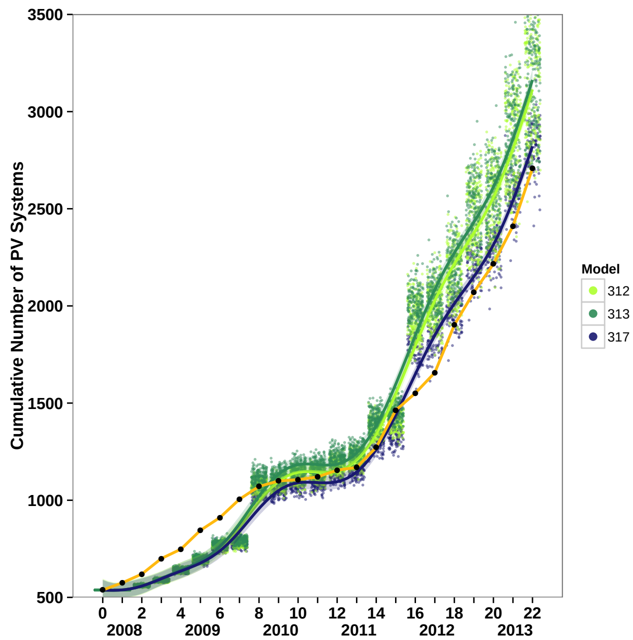


Figure 7.4: Informational dissemination campaign simulation targeting high betweenness agents.

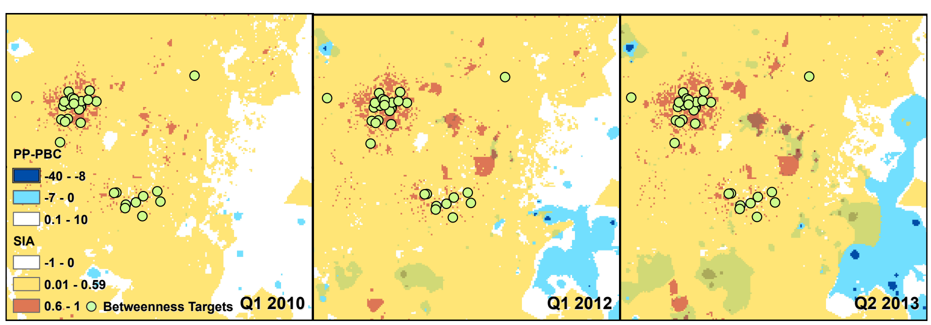


Figure 7.5: Evolution of attitudes and economic capacity under informational dissemination campaign scenario targeting high betweenness nodes.

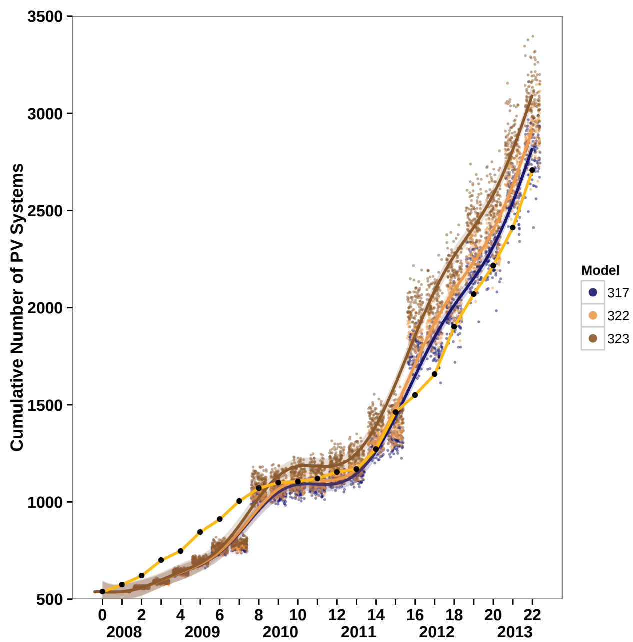


Figure 7.6: Informational dissemination campaign simulation targeting high closeness agents.

7.1.4 Eigenvector Centrality

Batches 324 and 325 look at an informational campaign targeted at 100 nodes with high Eigenvector centrality. Eigenvector centrality accounts for the centrality of an agent's neighbors in calculating its degree of influence. In this way, a low-degree node can be central if it is connected to a few high-degree nodes. However, this can have the effect of amplifying spatial clustering in geographic networks. As the map in figure 7.1 shows, agents with high eigenvector centrality are all concentrated in one high-density area in south Austin, and tend to be low-*pbc* agents. Thus, any visible effect from the informational campaign is slightly negative in terms of attitude as uncertainty is decreased, creating a slight local cooling effect. This is seen in figure 7.8.

7.2 Tiered Rebates

The potential to vary rebate amount to customers potentially addresses two areas of interest for policy makers. First, the fact that solar PV owners tend to be much wealthier than average has raised equity concerns in relation to rebates for solar PV. Because rebates are drawn from publicly generated revenues, they can have a regressive effect [91, 106]. Further, due to heterogeneity in wealth and price elasticity, when offered the same rebate, some consumers are capturing a surplus: for example, if the same consumer would have bought their system if given a $4,000 rebate, but received a $5,000 rebate, the extra $1,000 is surplus to the customer.

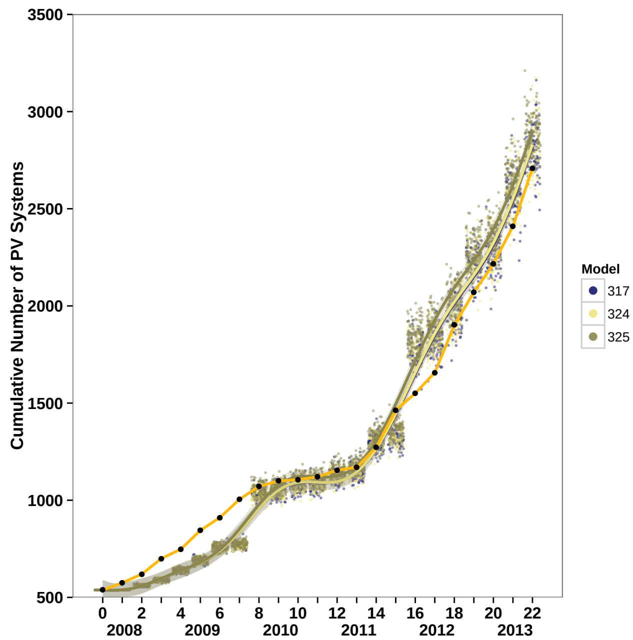


Figure 7.7: Informational dissemination campaign simulation targeting high Eigenvector centrality agents.

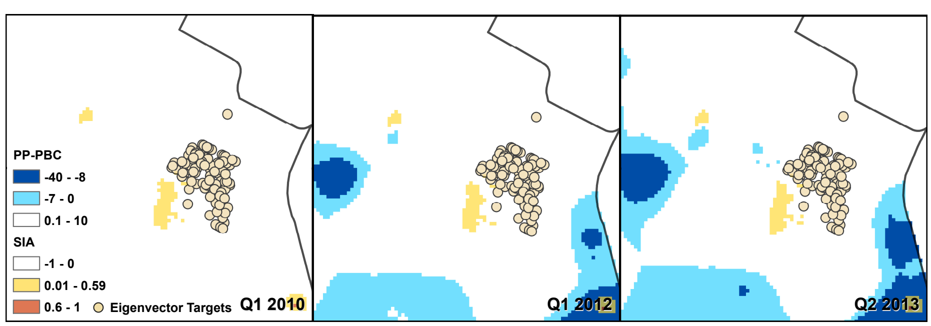
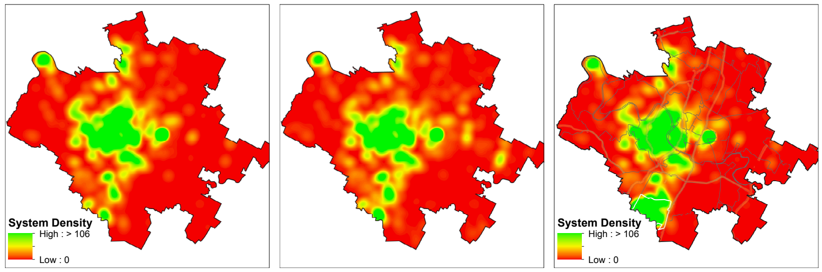


Figure 7.8: Evolution of attitudes and economic capacity under informational dissemination campaign scenario targeting high Eigenvector centrality nodes.

Second, due to transmission congestion, the occurrence of load-pockets and other factors [14], the electricity generated by solar PV systems may have different values for electric utilities based on location. The potential to offer different rebates in different locations based on value increases the ability of the utility to take advantage of solar energy and could allow for greater penetration levels without grid destabilization.

Therefore, we simulate a two-tiered utility offering: one at the empirical rate in dollars per Watt, and another at a higher rate. The first scenario (Batch 327) explores offering \$0.25 above the empirical rate to all customers in the bottom wealth (W) quartile. The impact of this offering over space can be seen in figure 7.9. The left panel (a) shows density in the base-case (Batch 317) for the selected zip code, the middle panel (b) shows Batch 327, where a tiered rebate (\$0.25) is offered to low wealth households, and the right panel (c) shows Batch 328, where the increased rebate is offered to all households in the target zip code. The density of installations increases in south and east Austin (generally low wealth areas), and decreases slightly in north and west Austin (generally high wealth areas).

In (Batch 328) we select just one zip code, intended to represent any potential locationally relevant segment (for example a dense load-pocket) and offer the higher rebate to all the agents in that zip code. As expected, this scenario results in a large increase in adoptions in that zip code relative to the base case. In the base case, on average 42 of the 5376 households in the zip code (0.78%) became PV adopters by Q2 2013. Under the tiered rebate, on average



(a) Base-case (b) Batch 327 (c) Batch 328

Figure 7.9: Spatial assessment of tiered rebate scenario simulations.

594 of the 5376 households (11.05%) became adopters. Next (Batch 308), we simulate a mixed strategy–attempting to stimulate adoption in one key area among low wealth households. Batch 309 repeats this with a slightly higher rebate ($0.50 above the empirical rate).

Tiered rebates have a large impact on the mean wealth distribution for adopters. Figure 7.13 plots the mean home value for adopters over time in the tiered rebate scenarios described above. Average home value for non-adopters in the study area ($267,965.80) is much lower than that of adopters in all scenarios, regardless of the time period. Interestingly, in all scenarios, the average wealth of solar PV adopters is increasing. This reflects an area of poor model fit, as the empirical data suggests average home values have decreased overall, although the trend is not smooth over time. This is likely due to the model generally over-predicting installation density in central Austin, where home values tend to be higher.

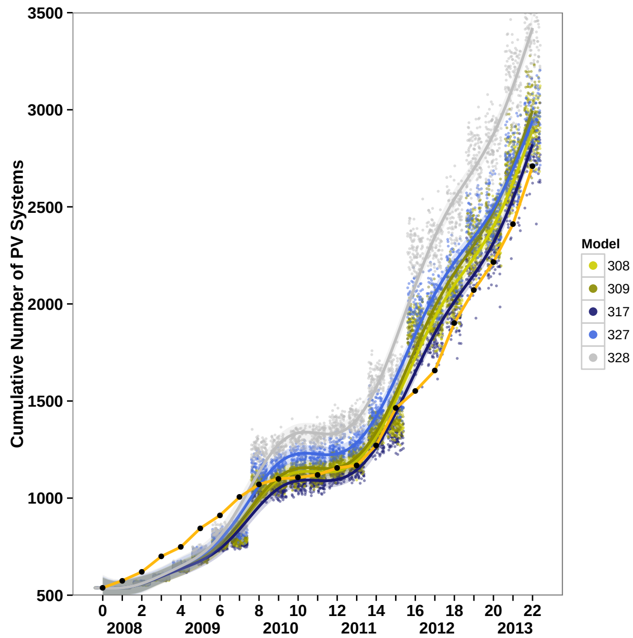
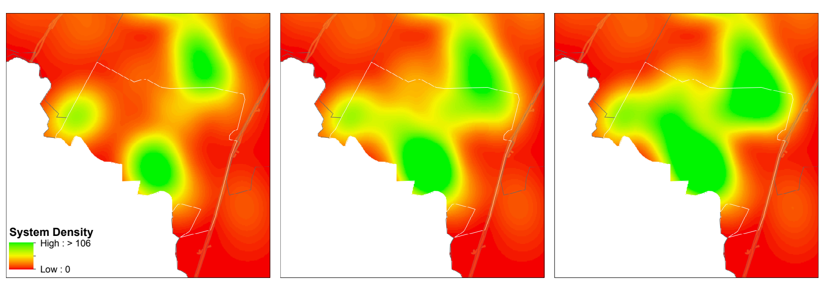


Figure 7.10: Cumulative number of installations over time in the Tiered Rebates Scenarios.



(a) Base-Case, Batch 317) (b) Tiered Rebate ($0.25), Batch 308 (c) Tiered Rebate ($0.50), Batch 309

Figure 7.11: Low income tiered rebate scenario impact on Q2 2013 system density, target zip code, highlighted by white dashed line.

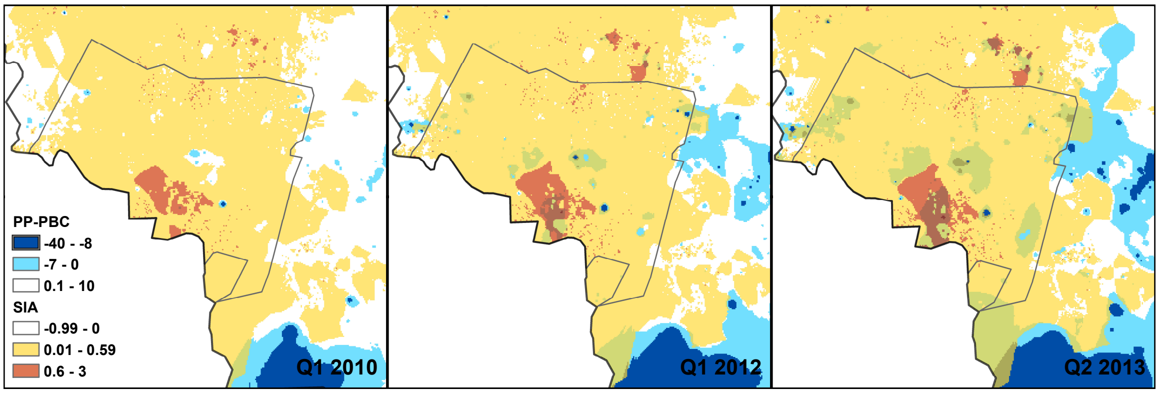


Figure 7.12: Evolution of *sia* and *pbc* in the tiered rebate scenario model 308.

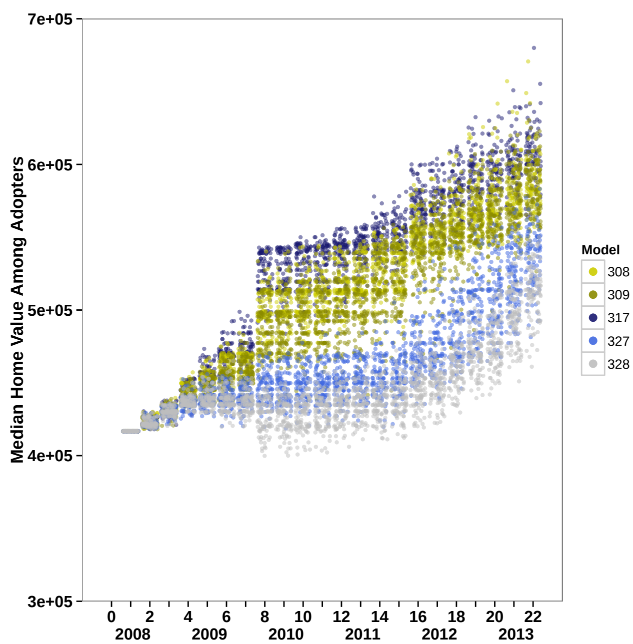


Figure 7.13: Mean wealth for adopters in the scenario simulations.

7.2.1 Alternative Rebate Schedules

The schedule of utility rebates has a large effect on the value proposition of solar PV. Further, there is the possibility of optimizing the rebate schedule in order to take advantage of peer effects on the demand side, and learning-by-doing on the supply side. In different market contexts, it may make sense to have a declining rebate schedule, or a flat rebate schedule. The empirical rebate schedule and the two simulated schedules used as scenarios are shown in figure 7.14. In both the flat and the steep scenario the average offering over the course of the study time-frame is held at the empirical average(2.51). The steep schedule starts at \$9.5 per Watt, and declines exponentially according to equation 7.1.

$$R_t = R_{t-1}^{0.9} \tag{7.1}$$

Altering the rebate schedule has a large impact on cumulative installations. As expected, the unconstrained steep schedule (Batch 330) shows very rapid installations, followed by a long period of gradual increase. The constrained steep case (Batch 334) shows a similar pattern, but the budget is quickly exhausted due to the size of the rebates being offered. This eventually results in fewer cumulative installations than the base-case. The unconstrained flat schedule (Batch 329) shows markedly fewer installations early, but bypasses the base-case in 2013. The flat constrained case (Batch 333) results in a few more systems than the base-case on average, trading off early sys-

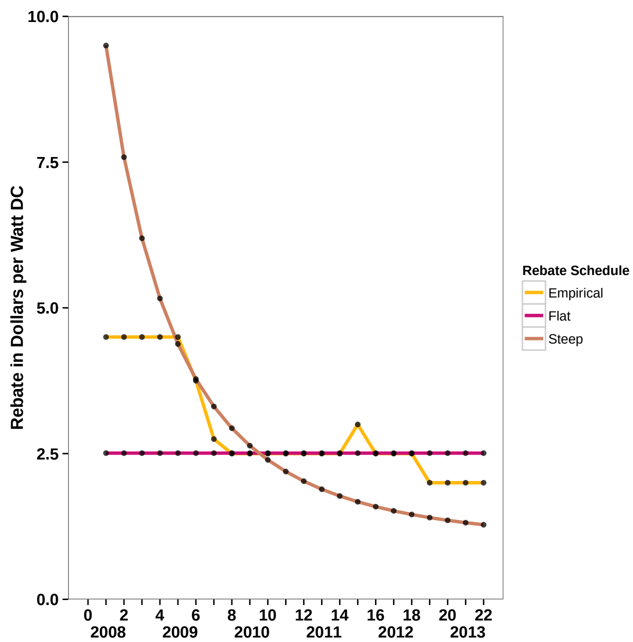
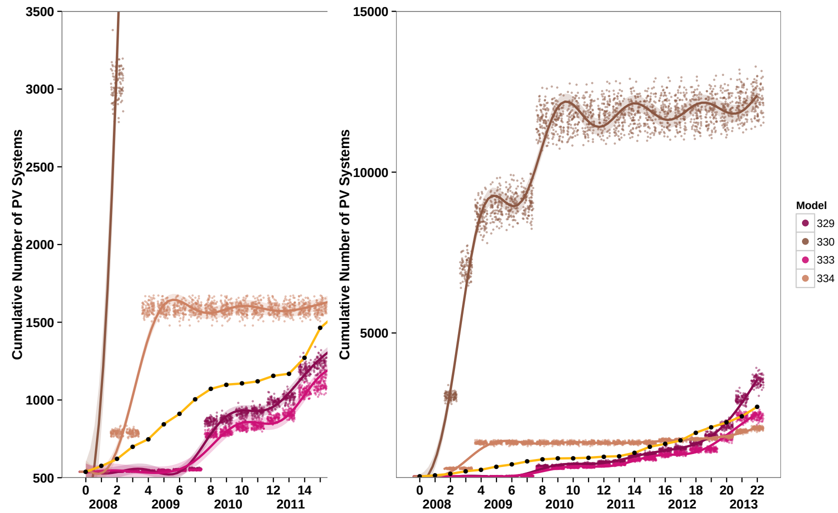


Figure 7.14: Rebate offerings by the electric utility agent over time in the empirical case and two hypothetical scenarios.



(a) a (b) a

Figure 7.15: Cumulative number of installations over time in the adjusted rebate schedule scenarios.

tems for late systems. Results are shown in table 7.2 and figure 7.15. The budget-constrained cases are much cheaper on a cost-per-system basis. Both for the constrained and unconstrained cases, the program costs (per-system) are higher for the flat rebate schedule, and much higher for the steep schedule relative to the base-cases.

Table 7.2: Alternative rebate schedule outcome summary.

Alternative Rebate Schedules				
Batch	% Increase	$ per Watt	$ per Watt Increase	Description
317	0	2.04	0.00	Base-case, un-constrained
330	335.27	4.56	2.52	Steep, un-constrained
329	25.04	2.25	0.21	Flat, un-constrained
326	0	2.19	0.00	Base-case, constrained
334	-14.19	3.26	1.07	Steep, constrained
333	1.06	2.22	0.03	Flat, constrained

Chapter 8

Discussion

This work has six major novel contributions to the scientific understanding of the diffusion of innovations and solar photovoltaics, each of which has a concrete policy application. These findings can be summarized as follows:

1. **Consumer Attitudes and Economics**

   In the popular discourse, the diffusion of solar PV, and the outcomes associated with it, such as the de-centralization and transformation of the electric power system [69], the threat to utility business models [14], large-scale emissions reductions [66, 112], and decreasing grid stability [39], is thought of only as a function of economics. However, many current solar PV owners are not wealthier than average and were able to purchase panels at a time when the systems were not highly profitable. This suggests that other factors are acting to limit solar diffusion. As shown in figure 4.7, the actual market penetration at any given point is dwarfed by the potential market defined by economic capability. Differences in attitude and information explain this gap. Importantly, both attitude and information are inherently dynamic, much more so than economic capacity: people change their minds, and become informed.

Addressing this information and attitude gap is a golden opportunity for policy-makers looking for low-cost ways to reduce greenhouse gas emissions, meet RPS requirements, achieve distributed generation targets, or increase power generation capacity.

2. **Successful Informational Dissemination Campaigns**

   While this golden opportunity certainly exists, the way outcomes can be influenced by information dissemination campaigns will be determined by the structure of the social network. Using influential individuals to disperse this information is a low cost way for utilities, solar installers, or non-profits to disperse accurate information to the public, more so than randomly located individuals. In geographically local small-world networks, individuals with high betweenness centrality are ideal for information dispersal due to their ability to act as informational brokers between neighborhoods that could otherwise be isolated. Further, in a market where prices are declining over time, because targeted informational campaigns increase the number of installations in early periods, they have the potential to increase costs for the utility above the cost of program implementation.

3. **Addressing Equity Issues**

   Ensuring that rebate dollars are not inequitably captured requires classifying households by wealth. Measures like income or home-value, while not perfect, can be acceptable proxies for wealth. Tiered rebate structures can change the wealth distribution of adopters, and decrease the magnitude of the wealth inequality. In the scenario offering \$0.25 to the bottom 25% of households by market home-value, the magnitude of the inequality was reduced by 22.6% on average. However, this will decrease the number of systems installed under a constrained-budget environment and increase program costs. Further, if a low-wealth area does not have a reasonable installed base of PV systems, the rebate offering will not have its full effect due to informational and attitudinal deficiencies. In low installed-base areas informational campaigns should be combined with tiered rebates.

4. **Increasing Value of Solar in the Grid**

   Simulations suggest that utilities can easily increase the installed base in a load-pocket through targeted rebates, if information is readily available through existing solar PV systems in the neighborhood. However, because the offering is associated with costs, utilities will need to perform granular value-of-solar calculations to find a suitable rebate offering. An ideal (but challenging and potentially costly) way this could be implemented is to perform value-of-solar calculations at the feeder level, and pay solar PV owners this rate for the electricity produced by their systems. We will attempt to better simulate this potential scenario in future work.

5. **Designing Rebate Schedules**

While optimization was not performed on the rebate schedule design, the path to doing so by minimizing the simulated $ per system metric in table 7.2 is clear. Clearly the major influencing factors are the system price curve and the system size curve, as the utility has costs in proportion to the number of Watts added to the grid. In smaller markets that act as price-takers, and thus are not able to generate learning-by-doing effects in suppliers, policy-makers should not expect to derive as much value from steeply declining rebate schedules. Further, it is noteworthy that on a per-system cost basis, the empirical utility rebate schedule outperformed the two hypothetical cases in the budget constrained as well as unconstrained case. Further, there is evidence that using hard budget constraints reduces program costs on a per-system basis, although it does not appear that this is the way the electric utility has operated in our current study area. Because attitudes and prices continue to change in the absence of rebate hand-outs, budget constraints may be an effective way for utilities to reduce program costs.

6. **Modeling solar PV diffusion** Distributed rooftop solar PV is an important part of the emerging portfolio of energy technologies that will shape the future electricity generation landscape. The ability to accurately model how much PV will be added to the grid and where will impact the important decisions made by utilities and governments regarding future investments, climate change, air quality and human health. Apart from the shortcomings addressed in chapter 1, traditional models

> (and even some ABMs) tend to inadequately address either the social, economic, or attitudinal sub-models that interact in consumer decision-making. The lack of empirical grounding in any of these components has large effects on simulation outcomes over time and space. Even fitted random models are not able to fully replicate existing patterns, and perform far worse on emergent pattern replication.

We hope to have demonstrated the value and potential of empirical agent-based modeling, particularly in the ability to generate hypothetical policy scenarios. Although other socio-technical systems were not explored at the level of detail that solar PV was, it is very likely that these general lessons apply. Emergent technologies, particularly those with large environmental benefits, will likely require policy measures to internalize market externalities and reach market penetration levels at which these benefits are realized. As both the granularity and availability of socio-technical data increase, the ability for modeling complex human systems will provide new opportunities for informed policy design. To make an analogy from the diffusion of innovations literature, simulation provides 'trialability,' which is absolutely critical in designing effective, equitable, and cost-efficient policies and programs.

8.1 Future Work

Work on the SECAD model is ongoing, and will focus on the development of further policy and decision-support scenarios. A few additional scenarios we plan to test include: electricity price change sensitivity, rebate schedule optimization, strategic seeding of systems in low-income neighborhoods, smart-targeting (proving information to low-attitude households, and economic incentives to low-wealth households), and induced peer effects through social-media interventions. Further, the fitted model allows us to test other important hypotheses that are difficult to assess in the field, such as the relative importance of strong and weak ties in the formation of new adoption clusters, and time-to-economic versus time to attitudinal activation models. We also intend to examine the role of peer-effects in our model in parallel with existing survey-based research [118].

Beyond these scenarios, we plan on expanding the SECAD model to include supply-side agents (installers, contractors, etc.), who use various marketing techniques to push technologies into the marketplace. In addition to creating more realistic agent interactions, this will allow for the testing and assessment of competing business strategies within a dynamic marketplace. We would also like to expand the model to different residential markets, and are continuing to collect data from different municipalities toward this goal.

Network topology plays a large role in adoption outcomes. To improve the extra-local connections in the small world network, we must determine the structure of relevant non-neighborhood connections influential to decision-making about solar PV. Future work will seek to empirically derive these non-local connections, or at least reduce the degree to which they are random. We are working on a nation-wide survey to reveal these connections that is scheduled to be deployed sometime this year.

Appendices

Appendix A

Supporting Information

A.1 Relevant Survey Questions

The UT solar survey data used in this analysis were gathered in two phases: the first in 2011, the second in 2012, and the third in 2013. The survey administered electronically (online) was used to construct a novel dataset of solar PV adopters in Texas. The majority of respondents were located in the greater Austin and Dallas-Fort Worth metro areas. The sample of complete responses for the Austin area (616) reflects a total response rate of (22.5%).

The following questions were used in this analysis (question numbers refer to survey ordering):

1. 2.2 How important were the following factors in your final decision to install a PV system (1-5)?

    Your general interest in energy and electricity generation

    Your evaluation that solar PV is a good financial investment

    Reducing impact on the environment by using a renewable energy source

Influence of others in the neighborhood with PV systems

Influence of a close acquaintance not from your neighborhood

2. 3.1 Please characterize the overall profitability of your PV system at the time you decided to install.

3. 3.4 What resulting values from the above financial estimates (Net Present Value (NPV); Rate of Return; Payback Period; Net Monthly Savings) did you arrive at?

4. 4.3 How many other owners of a PV system did you have contact with regarding PV before installing your system?

   How many of these contacts were in your neighborhood?

5. 4.5 As far as you know, how many PV systems were in your neighbrohood when you were deciding to install?

6. 4.6 How much do you agree or disagree with each of the following statements about PV systems in your neighborhood during the decision-making process (1-5)?

   PV Systems in the neighborhood motivated me to seriously consider installing one

   Seeing other PV systems in the neighborhood gave me confidence that it would be a good decision to install one

7. 6.2 In general, how concerned were you about environmental issues before

the installation of your PV system?

8. 6.3 How willing are you to pay much higher prices in order to protect the environment?

9. 6.4 How would you describe the level of concern for environmental issues in your neighborhood (a radius of roughly 5-10 blocks around your house) (1-5)?

A.2 Diagnostics for Regression (equation 5.5)

The model generates expected values for socially informed attitude about PV through correlated regressors: wealth (W), size of the lot in square feet (s^P), and tree cover in square feet (T). Four data points were found to have a large amount of influence in the regression (Cook's D ¿ 0.2), and were removed. Four of the plots used to identify these points are shown in figure A.1. Due to the large number of potential covariates, model selection was performed via a stepwise procedure using AIC and $AdjR^2$. The model presented in equation 5.5 is the result of this process, having an $AdjR^2$ of 0.21, and an AIC of 557.44. The p-value for the model is <0.0001. Further, analysis of residuals suggests good fit. Figure A.2 shows four additional diagnostic plots: the residuals appear to be randomly and normally distributed, and there are no highly influential points after those shown in figure A.1. Surprisingly, while sP was consistently a significant predictor, in none of the models was s (size of the home in square feet) significant.

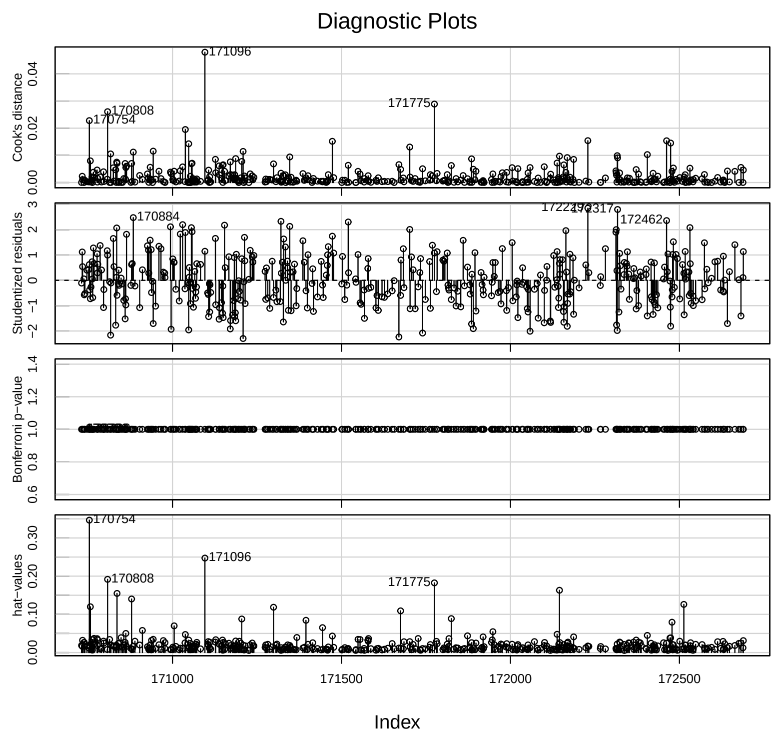


Figure A.1: Cooks distance, standardized residuals, Bonferroni's P-values, and hat-matrix values for the model were plotted to identify influential observations. Four of these were removed.

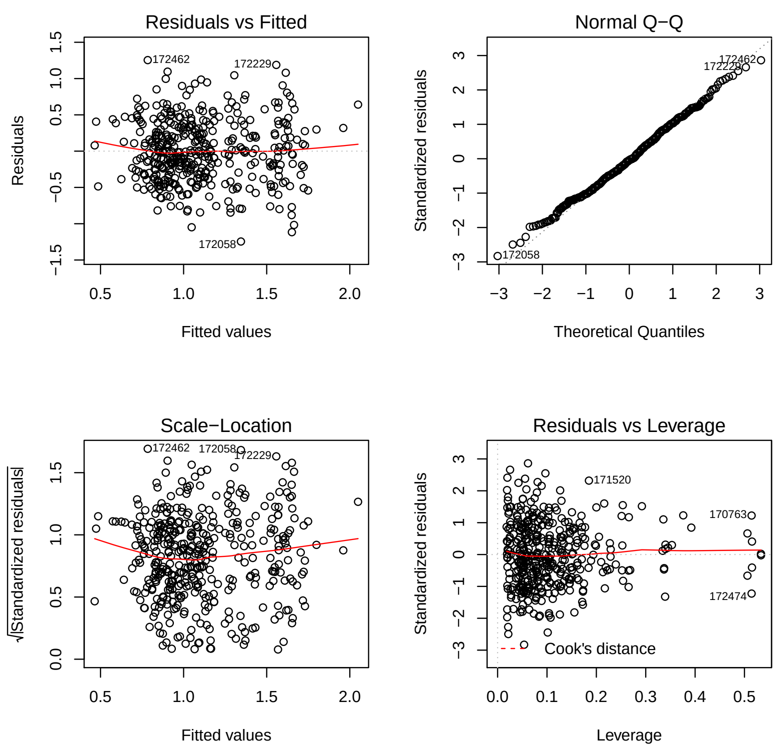


Figure A.2: Diagnositic plots for equation 5.5

Appendix B

Sensitivity Testing

Because agent-based modeling involves multiple parameters (see table 4.3), it is important to test the model's sensitivity to each model parameter across a range of potential values. Sensitivity testing serves five distinct purposes.

1. Verify model structure and parameter definitions. From a verification standpoint, sensitivity testing shows that each parameter is behaving as expected, or if not, shows which parameters require further scrutiny.

2. Increase understanding of the model. By altering parameter values, it is possible to quantify each parameter's marginal effect. This is useful in allowing the parameter to be understood in a classical econometric sense. Further, full parameter sweeps can generate insight into parameter interaction effects.

3. Demonstrate ABM flexibility and range of outcomes. By varying parameters along realistic ranges, the total range of outcomes that the model will generate can be found. This can be compared to the range generated by each parameter to gain an understanding of relative sensitivity.

4. Cast insight on the system being modeled. Through identifying marginal

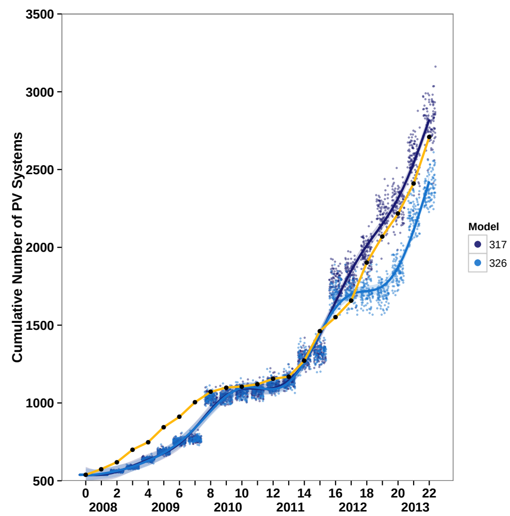
impacts, and potentially by finding unforeseen interaction effects, it is possible to generate inference into the system being modeled. This is only possible given an empirical framework, and is especially relevant to policy-makers wary of unintended consequences.

5. Show potential alternative scenarios. While the generalization of any model results should be approached with extreme caution, altering the ABM parameters can potentially shed light into alternative markets, or at least show the effect of a range of behavioral nudges.

B.1 Budget

The model is sensitive to the presence of a budget constraint, implemented in model 326. In model number 317, we remove the constraint, simulating the utility's willing to take on an deficit. This alters the shape of curve by effectively capping installations until the next fiscal year if the budget is fully allocated, as is shown in figure B.1. Here we show model sensitivity results for constrained and unconstrained models. Importantly, the budget-constrained cases were run under slightly different parameters than batch 326, which is why the curves appear altered relative to those seen in chapter 7. However, the effects remain the same.

Figure B.1: Model sensitivity of the utility budget constraint



B.2 Intensity of Interaction

Figure B.2 shows the model's sensitivity to μ, which controls the speed of convergence in the model. The budget constraint was applied in each of these models. μ does not have a linear relationship with cumulative adoption, because with very quick convergence, average sia levels in the model are repressed by influence of non-adopters. Because uncertainty U is also dynamic, if attitudes stay low they become more difficult to influence over time. Table B.1 lists RMSE statistics for this parameter sensitivity. Low μ minimizes marginal RMSE, while a moderate μ of 0.5 minimizes cumulative RMSE.

The scenario's showing the sensitivity to μ without the budget constraint are shown in figure B.3. The results are similar in that μ has a non-linear effect on the number of cumulative installations.

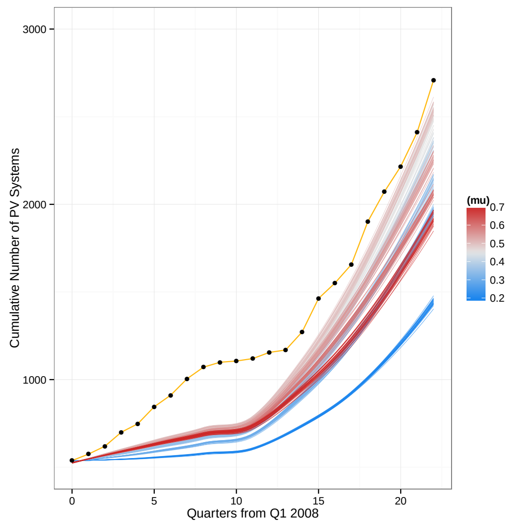


Figure B.2: Model sensitivity to the weight on each interaction, μ.

Table B.1: Marginal and cumulative RMSE for μ sensitivity.

	RMSE		
Batch	Marginal Adoption	Cumulative Adoption	μ
214	95.73	639.21	0.2
218	128.78	431.88	0.3
217	139.18	359.90	0.35
215	152.20	295.67	0.4
216	161.60	262.14	0.45
219	207.86	242.50	0.5
220	147.18	297.50	0.55
221	134.93	355.25	0.6
222	112.22	414.18	0.7

Table B.2: Means of spatial error statistics for μ sensitivity.

Spatial Error, μ				
Batch	$b_i - a_i$	Correlation, r	Fuzzy Numerical, Ξ	
218	1.73	4.91	0.35	0.3
216	0.62	4.84	0.40	0.45
222	1.64	3.89	0.40	0.7

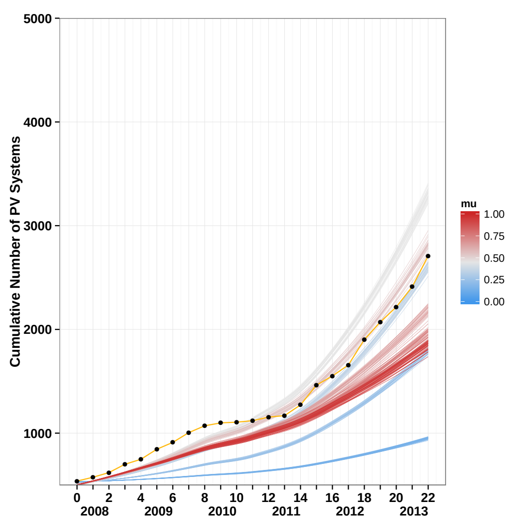


Figure B.3: Model sensitivity to the weight on each interaction, μ, in unconstrained case.

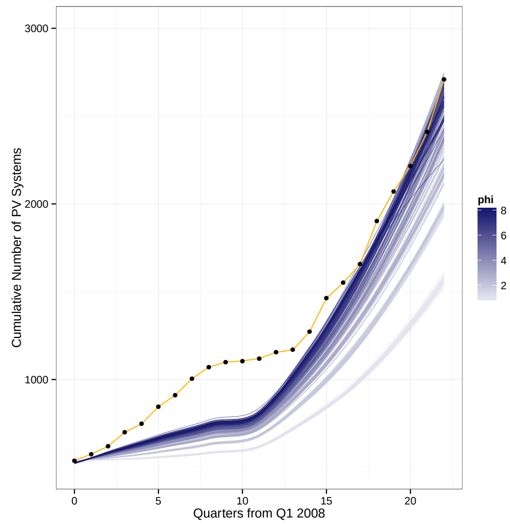


Figure B.4: Model sensitivity to number of interactions, ϕ, with budget constraint.

B.3 Number of Interactions

Increasing the number of interactions (ϕ)that each agent has per quarter has a positive marginal effect on the number of cumulative installations, as shown in figure B.4. The budget constraint was applied in each of these models. Table B.3 lists RMSE for sensitivity to ϕ. Low ϕ minimizes marginal error while high ϕ of 7 minimizes cumulative error in the budget constrained case.

After lifting the budget constraint, increasing the number of interactions has a larger effect on the number of cumulative installations (figure B.5), however, there are diminishing marginal returns–increasing from one to two has a much larger effect than increasing from seven to eight.

Table B.3: Marginal and cumulative RMSE for ϕ sensitivity.

RMSE			
Batch	Marginal Adoption	Cumulative Adoption	ϕ
227	100.28	588.58	1
226	130.57	429.05	2
225	143.65	347.09	3
226	130.57	329.05	4
216	161.60	262.14	5
228	168.41	240.95	6
229	172.49	229.08	7
230	173.87	229.19	8

Table B.4: Means of spatial error statistics for ϕ sensitivity.

Spatial Error, ϕ				
Batch	$b_i - a_i$	$abs(b_i - a_i)$	Fuzzy Numerical (Ξ)	ϕ
227	1.99	4.81	0.35	1
216	0.62	4.84	0.40	5
230	0.49	4.68	0.41	8

Table B.5: Marginal and cumulative RMSE for ϕ sensitivity, without the budget constraint.

RMSE			
Batch	Marginal Adoption	Cumulative Adoption	ϕ
267	90.76	646.22	1
268	103.72	366.64	2
269	140.94	290.12	3
270	162.10	380.32	4
250	188.76	432.09	5
271	206.22	506.63	6
272	220.34	565.35	7
273	231.89	612.78	8

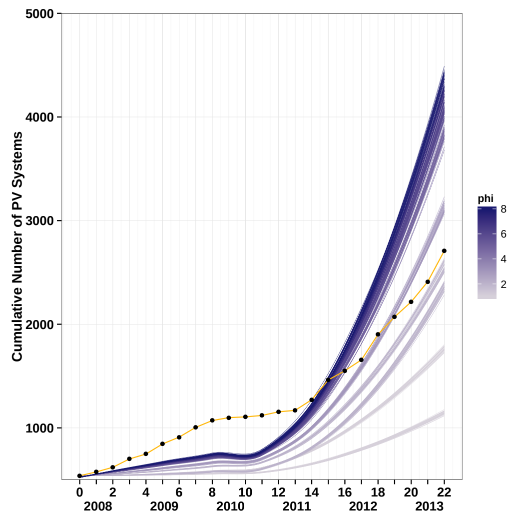


Figure B.5: Model sensitivity to number of interactions, ϕ, without the budget constraint.

B.4 Random Connections

The proportion of random connections (λ^r) in the agent social networks has an interesting effect on diffusion outcomes over time. As shown in figure B.6, more random networks show increased adoption over empirically derived networks until $t11$ (Q3 2010), when the effect is inverted. This is due to the trade-off between the speed of information spread and the strength of feedback effects. Early, static geographic networks help reinforce positive attitudes toward solar PV, leading to high cumulative installations. In the language of TPB, these local effects create a strong social norm. However, the same local structure creates barriers to widespread adoption as information disseminates more slowly compared to the random networks ($\lambda = 1$), where information can take rapid shortcuts across the entire graph. The budget constraint was

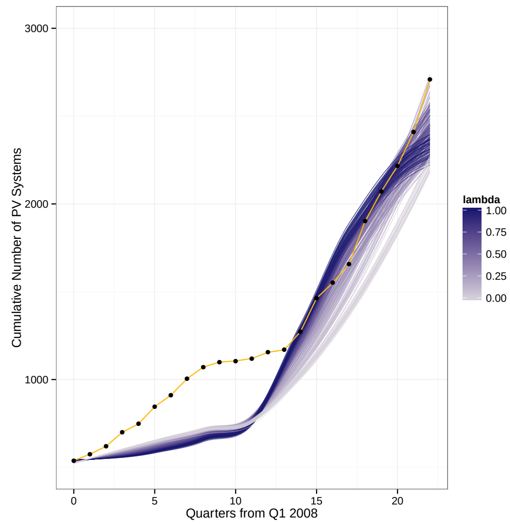


Figure B.6: Model sensitivity to the proportion of random connections, λ.

applied in each of these models. Table B.6 lists λ^r sensitivity RMSE. The base-case proportion of random connections, 0.1, minimizes marginal error, while the lowest cumulative error is seen with 20% random connections.

After lifting the budget constraint, the same effect is observed, and can be seen more clearly for high λ^r simulations. The low λ^r effect of increasing adoptions early on is still visible, although somewhat less-so.

B.5 Perceived Behavioral Control Ceiling

Altering the scale ceiling for *pbc* changes the way the agent's calculated perceived behavioral control (equation 5.11) is mapped onto a scale representative of the payback period. Increasing this value increases the number of agents that can potentially adopt solar at a given time t, which has the logical conse-

Table B.6: Marginal and cumulative RMSE for λ^r sensitivity.

RMSE			
Batch	Marginal Adoption	Cumulative Adoption	λ^r
240	142.96	311.80	0.0
216	128.78	262.14	0.1
231	176.93	247.01	0.2
232	173.11	265.47	0.3
233	182.85	260.77	0.4
234	192.12	261.67	0.5
235	204.03	266.66	0.6
236	212.34	284.08	0.7
237	219.61	301.97	0.8
238	224.92	310.69	0.9
239	231.88	318.20	1.0

Table B.7: Means of spatial error statistics for λ^r sensitivity.

Spatial Error, λ^r				
Batch	$b_i - a_i$	$abs(b_i - a_i)$	Fuzzy Numerical (Ξ)	λ^r
240	1.22	5.09	0.38	0.0
216	0.62	4.84	0.40	0.1
233	1.13	3.78	0.43	0.4
234	0.97	3.73	0.44	0.5
239	1.19	3.37	0.37	1.0

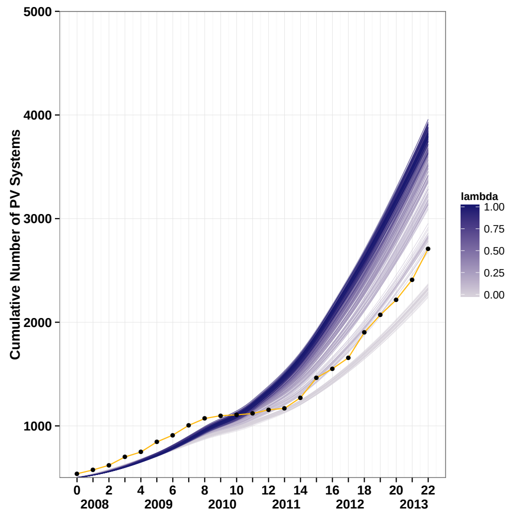


Figure B.7: Model sensitivity to the proportion of random connections, λ without the budget constraint.

quence of increasing the cumulative number of installations predicted by the simulation. The budget constraint was applied in each of these models. Temporal error statistics are shown in table B.8. The lowest marginal RMSE is observed with the 28.56 floor, while the lowest cumulative RMSE is seen with the 24.56 floor. Unconstrained results are shown in figure B.9, and show the same relationship as the constrained case.

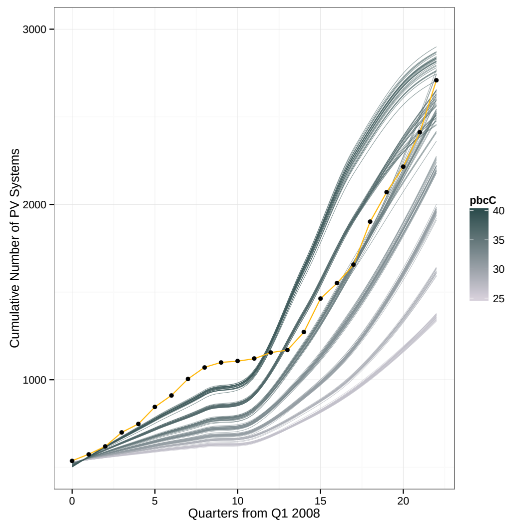


Figure B.8: Model sensitivity to the scale ceiling for mapping *pbc* to payback, *pbcC*.

Table B.8: Marginal and cumulative RMSE for *pbcC* sensitivity.

RMSE, *pbcC*			
Batch	Marginal Adoption	Cumulative Adoption	*pbcC*
243	98.09	635.86	26.56
242	95.77	551.85	28.56
241	116.26	432.37	30.56
216	161.60	262.14	32.56
244	175.32	223.66	34.56
245	207.86	242.50	36.56
246	255.86	394.49	38.56

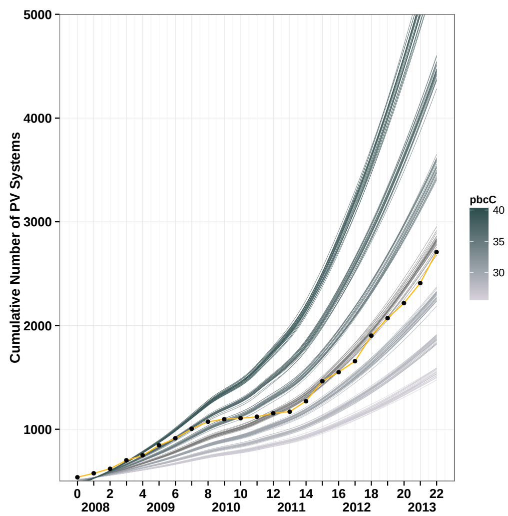


Figure B.9: Model sensitivity to the scale ceiling for mapping *pbc* to payback, *pbcC*, without the budget constraint.

Table B.9: Means of spatial error statistics for *pbcC* several sensitivities.

Spatial Error, *pbcC*				
Batch	$b_i - a_i$	$abs(b_i - a_i)$	Fuzzy Numerical (Ξ)	$pbcC$
243	2.58	4.41	0.33	26.56
216	0.62	4.84	0.40	32.56
246	0.15	5.85	0.38	38.56

Bibliography

[1] L. Adamic. The small world web. In Serge Abiteboul and Anne-Marie Vercoustre, editors, *Research and Advanced Technology for Digital Libraries*, volume 1696 of *Lecture Notes in Computer Science*, pages 443–452. Springer Berlin Heidelberg, 1999. ISBN 978-3-540-66558-8. doi: 10.1007/3-540-48155-9_27. URL `http://dx.doi.org/10.1007/3-540-48155-9_27`.

[2] Icek Ajzen. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211, 1991. ISSN 0749-5978.

[3] Icek Ajzen. Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, 32(4):665–683, 2002. ISSN 1559-1816. URL `http://dx.doi.org/10.1111/j.1559-1816.2002.tb00236.x`.

[4] S.J. Alam and A. Geller. Networks in agent-based social simulation. In A.J.: Heppenstall, A.T.; Crooks, L.M.; See, and M. Batty, editors, *Agent-Based Models of Geographical Systems*, pages 199–216. Springer, 2012.

[5] J. Alderman, D. McCollin, S.A. Hinsley, P.E. Bellamy, P. Picton, and R. Crockett. Modelling the effects of dispersal and landscape configu-

ration on population distribution and viability in fragmented habitat. *Landscape Ecology*, 20(7):857–870, 2005.

[6] Christopher J. Armitage and Mark Conner. Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, 40(4):471–499, 2001. ISSN 2044-8309. URL http://dx.doi.org/10.1348/014466601164939.

[7] Nava Ashraf, B. Kelsey Jack, and Emir Kamenica. Information and subsidies: Complements or substitutes? *Journal of Economic Behavior & Organization*, 88(0):133 – 139, 2013. ISSN 0167-2681. doi: http://dx.doi.org/10.1016/j.jebo.2012.12.031. URL http://www.sciencedirect.com/science/article/pii/S0167268112003034. Asian Institutional Economics.

[8] T. Baker, S. Miller, S. Prager, and D. Legg. Disaggregating human population for improved land use management in kenya. *Journal of Land Use Science*, 5(4):237–257, 2010. doi: 10.1080/1747423X.2010.500685. URL http://www.tandfonline.com/doi/abs/10.1080/1747423X.2010.500685.

[9] S. Balietti. *Social Self-Organization: Understanding Complex Systems*, chapter Agent-Based Modeling, pages 25–70. Springer, 2012.

[10] Frank M. Bass. A simultaneous equation regression study of advertising and sales of cigarettes. *Journal of Marketing Research*, 6(3):291–300, 1969. ISSN 00222437. URL http://www.jstor.org/stable/3150135.

[11] Giuliana Battisti and Paul Stoneman. The intra-firm diffusion of new process technologies. *International Journal of Industrial Organization*, 23(1-2):1–22, 2005. ISSN 0167-7187. doi: http://dx.doi.org/10.1016/j.ijindorg.2004.12.002. URL `http://www.sciencedirect.com/science/article/pii/S0167718705000081`.

[12] Thomas Berger. Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. *Agricultural Economics*, 25(2-3):245–260, 2001. ISSN 1574-0862. doi: 10.1111/j.1574-0862.2001.tb00205.x. URL `http://dx.doi.org/10.1111/j.1574-0862.2001.tb00205.x`.

[13] S. C. Bhargava, Arun Kumar, and A. Mukherjee. A stochastic cellular automata model of innovation diffusion. *Technological Forecasting and Social Change*, 44(1):87–97, 8// 1993. ISSN 0040-1625. URL `http://www.sciencedirect.com/science/article/pii/004016259390008U`.

[14] Griselda Blackburn, Clare Magee, and Varun Rai. Solar valuation and the modern utility's expansion into distributed generation. *The Electricity Journal*, 2014.

[15] M. Boguna and R. Pastor-Satorras. Epidemic spreading in correlated complex networks. *Physical Review E*, 66:1–4, Oct 2002. doi: 10.1103/PhysRevE.66.047104. URL `http://link.aps.org/doi/10.1103/PhysRevE.66.047104`.

[16] J.D. Bohlmann, R.J. Calantone, and M. Zhao. The effects of market

network heterogeneity on innovation diffusion: An agent-based modeling approach. *Journal of Product Innovation Management*, 27(5):741–760, 2010. ISSN 1540-5885. doi: 10.1111/j.1540-5885.2010.00748.x. URL http://dx.doi.org/10.1111/j.1540-5885.2010.00748.x.

[17] Eric Bonabeau. Agent-based modeling: Method and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99:7287–7287, 2002.

[18] P. Bonacich. Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182, 1987. ISSN 00029602. URL http://www.jstor.org/stable/2780000.

[19] P. Bonacich. Some unique properties of eigenvector centrality. *Social Networks*, 29(4):555 – 564, 2007. ISSN 0378-8733. doi: http://dx.doi.org/10.1016/j.socnet.2007.04.002. URL http://www.sciencedirect.com/science/article/pii/S0378873307000342.

[20] W.M. Briggs and R.A. Levine. Wavelets and field forecast verification. *Monthly Weather Review*, 125(6):1329, 1997. ISSN 00270644. URL http://ezproxy.lib.utexas.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a2h&AN=5654649&site=ehost-live.

[21] Michel Callon. *Society in the making: the study of technology as a tool for sociological analysis*, volume 550, chapter 4, pages 83–103. MIT press Cambridge, MA, 1987.

[22] Simona Cantono and Gerald Silverberg. A percolation model of eco-innovation diffusion: The relationship between diffusion, learning economies and subsidies. *Technological Forecasting and Social Change*, 76(4):487 – 496, 2009. ISSN 0040-1625. doi: http://dx.doi.org/10.1016/j.techfore.2008.04.010. URL `http://www.sciencedirect.com/science/article/pii/S0040162508000863`. Evolutionary Methodologies for Analyzing Environmental Innovations and the Implications for Environmental Policy.

[23] P.J. Carrington, J. Scott, and S. Wasserman. *Models and methods in social network analysis*. Cambridge university press, 2005.

[24] B. Casati, G. Ross, and D. B. Stephenson. A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorological Applications*, 11(2):141–154, 2004. ISSN 1469-8080. doi: 10.1017/S1350482704001239. URL `http://dx.doi.org/10.1017/S1350482704001239`.

[25] H. Choi, S.H. Kim, and J. Lee. Role of network structure and network effects in diffusion of innovations. *Industrial Marketing Management*, 39(1):170 – 177, 2010. ISSN 0019-8501. doi: http://dx.doi.org/10.1016/j.indmarman.2008.08.006. URL `http://www.sciencedirect.com/science/article/pii/S0019850108001557`. Case Study Research in Industrial Marketing.

[26] J.K. Clark, D.T. Wegener, and L.R. Fabrigar. Attitudinal ambivalence

and message-based persuasion: Motivated processing of proattitudinal information and avoidance of counterattitudinal information. *Personality and Social Psychology Bulletin*, 34(4):565–577, 2008.

[27] Nick Collier. Repast: An extensible framework for agent simulation. *The University of Chicago's Social Science Research*, 36, 2003.

[28] A.T Crooks and C.J.E Castle. *Agent-based models of geographical systems*, chapter The integration of agent-based modelling and geographical information for geospatial simulation, pages 219–252. Springer Dordrecht Heigelberg, 2012.

[29] A.T. Crooks and A.J. Heppenstall. Introduction to agent-based modelling. *Agent-based models of geographical systems*, pages 85–105, 2012.

[30] G. Deffuant and G Weisbuch. Probability distribution dynamics explaining agent model convergence to extremism. *Social Simulation: Technologies, Advances and New Discoveries*, pages 43–60, 2008.

[31] G. Deffuant, F. Amblard, G. Weisbuch, and T. Faure. How can extremism prevail? a study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation*, 5(4), 2002.

[32] G. Deffuant, S. Huet, J.P. Bousset, J. Henriot, G. Amon, G. Weisbuch, and M.A. Janssen. Agent-based simulation of organic farming conversion in allier département. *Complexity and ecosystem management: The theory and practice of multi-agent systems*, pages 158–187, 2002.

[33] G. Deffuant, S. Huet, and F. Amblard. An individual based model of innovation diffusion: Mixing social value and individual benefit. *American Journal of Sociology*, 110(4):1041–1069, 2005. ISSN 00029602. URL http://www.jstor.org/stable/10.1086/430220.

[34] Guillaume Deffuant, David Neau, Frederic Amblard, and Grard Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(1):87–98, 2000. URL http://www.worldscientific.com/doi/abs/10.1142/S0219525900000078.

[35] S.A. Delre, W. Jager, T.H.A. Bijmolt, and M.A. Janssen. Targeting and timing promotional activities: An agent-based model for the takeoff of new products. *Journal of business research*, 60(8):826–835, 2007.

[36] F. Deroian. Formation of social networks and diffusion of innovations. *Research Policy*, 31(5):835 – 846, 2002. ISSN 0048-7333. doi: http://dx.doi.org/10.1016/S0048-7333(01)00147-0. URL http://www.sciencedirect.com/science/article/pii/S0048733301001470.

[37] Philip M. Dixon. Ripley's k function. In *Encyclopedia of Environmetrics*. John Wiley & Sons, Ltd, 2006. ISBN 9780470057339. URL http://dx.doi.org/10.1002/9780470057339.var046.

[38] R. Ducrot, C. Le Page, P. Bommel, and M. Kuper. Articulating land and water dynamics with urbanization: an attempt to model natural resources management at the urban edge. *Computers, Environment and Urban Systems*, 28(12):85 – 106, 2004.

ISSN 0198-9715. doi: http://dx.doi.org/10.1016/S0198-9715(02)00066-2. URL http://www.sciencedirect.com/science/article/pii/S0198971502000662. ¡ce:title¿Geosimulation¡/ce:title¿.

[39] M.A. Eltawil and Z. Zhao. Grid-connected photovoltaic power systems: Technical and potential problemsa review. *Renewable and Sustainable Energy Reviews*, 14(1):112–129, 2010.

[40] Joshua M. Epstein. Why model? *Journal of Artificial Societies and Social Simulation*, 11(4):12, 2008. ISSN 1460-7425. URL http://jasss.soc.surrey.ac.uk/11/4/12.html.

[41] John E Ettlie. Adequacy of stage models for decisions on adoption of innovation. *Psychological Reports*, 46(3):991–995, 1980.

[42] T. P. Evans and H. Kelley. Multi-scale analysis of a household level agent-based model of landcover change. *Journal of Environmental Management*, 72(1-2):57 – 72, 2004. ISSN 0301-4797. doi: http://dx.doi.org/10.1016/j.jenvman.2004.02.008. URL http://www.sciencedirect.com/science/article/pii/S0301479704000787.

[43] T. Filatova, P. Verburg, D. Parker, and C. Stannard. Spatial agent-based models for socio-ecological systems: Challenges and prospects. *Environmental Modelling & Software*, 45(0):1 – 7, 2013. ISSN 1364-8152. doi: http://dx.doi.org/10.1016/j.envsoft.2013.03.017. URL http://www.sciencedirect.com/science/article/pii/S1364815213000807.

[44] D.W. Franks, J. Noble, P. Kaufmann, and S. Stagl. Extremism propagation in social networks with hubs. *Adaptive Behavior*, 16(4):264–274, 2008. doi: 10.1177/1059712308090536. URL `http://adb.sagepub.com/content/16/4/264.abstract`.

[45] C. Freeman. The national system of innovationin historical perspective. *Cambridge Journal of economics*, 19(1):5–24, 1995.

[46] L.C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215 – 239, 1978. ISSN 0378-8733. doi: http://dx.doi.org/10.1016/0378-8733(78)90021-7. URL `http://www.sciencedirect.com/science/article/pii/0378873378900217`.

[47] H. Gatignon and T.S. Robertson. Technology diffusion: an empirical test of competitive effects. *The Journal of Marketing*, pages 35–49, 1989.

[48] P.A. Geroski. Models of technology diffusion. *Research policy*, 29(4):603–625, 2000. ISSN 0048-7333.

[49] Arthur Getis and J Keith Ord. The analysis of spatial association by use of distance statistics. *Geographical analysis*, 24(3):189–206, 1992. ISSN 1538-4632.

[50] N. Gilbert and S. Bankes. Platforms and methods for agent-based modeling. *Proceedings of the National Academy of Sciences*, 99:7197–7198, 2002.

[51] A. N. Giovanis and C. H. Skiadas. A stochastic logistic innovation dif-

fusion model studying the electricity consumption in greece and the united states. *Technological Forecasting and Social Change*, 61(3):235–246, 7// 1999. ISSN 0040-1625. URL `http://www.sciencedirect.com/science/article/pii/S0040162599000050`.

[52] L.R. Glasman and D. Albarracan. Forming attitudes that predict future behavior: A meta-analysis of the attitude-behavior relation. *Psychological bulletin*, 132(5):778, 2006. ISSN 1939-1455.

[53] Jacob Goldenberg, Barak Libai, and Eitan Muller. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review*, 9(3):1–18, 2001.

[54] Pierre Goovaerts. *Geostatistics for natural resources evaluation*. Oxford university press, 1997.

[55] Zvi Griliches. Hybrid corn: An exploration in the economics of technological change. *Econometrica, Journal of the Econometric Society*, pages 501–522, 1957.

[56] V. Grimm, E. Revilla, U. Berger, F. Jeltsch, W. M. Mooj, S. F. Railsback, H. H. Thulke, J. Weiner, T. Weigand, and D. DeAngelis. Pattern-oriented mdeling of agent-based complex systems: lessons from ecology. *Science*, 310:987–991, 2005.

[57] Veronica Guerrieri, Robert Shimer, and Randall Wright. Adverse se-

lection in competitive search equilibrium. *Econometrica*, 78(6):1823–1862, 2010. ISSN 1468-0262. doi: 10.3982/ECTA8535. URL http://dx.doi.org/10.3982/ECTA8535.

[58] M. Gunther, C. Stummer, L. Wakolbinger, and M. Wildpaner. An agent-based simulation approach for the new product diffusion of a mvel biomass fuel. *J. Oper. Res. Soc.*, 62:12–20, 20011.

[59] A. Hagen. Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science*, 17(3):235–249, 2003. doi: 10.1080/13658810210157822. URL http://www.tandfonline.com/doi/abs/10.1080/13658810210157822.

[60] A. Hagen, B. Straatman, and I. Uljee. Further developments of a fuzzy set map comparison approach. *International Journal of Geographical Information Science*, 19(7):769–785, 2005. doi: 10.1080/13658810500072137. URL http://www.tandfonline.com/doi/abs/10.1080/13658810500072137.

[61] Bronwyn H Hall. Innovation and diffusion. Technical report, National Bureau of Economic Research, 2004.

[62] WesleyR. Hartmann, Puneet Manchanda, Harikesh Nair, Matthew Bothner, Peter Dodds, David Godes, Kartik Hosanagar, and Catherine Tucker. Modeling social interactions: Identification, empirical methods and policy implications. *Marketing Letters*, 19(3-4):287–304, 2008. ISSN

0923-0645. doi: 10.1007/s11002-008-9048-z. URL `http://dx.doi.org/10.1007/s11002-008-9048-z`.

[63] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3):2, 2002.

[64] Alison Heppenstall, Andrew Evans, and Mark Birkin. Using hybrid agent-based systems to model spatially-influenced retail markets. *Journal of Artificial Societies and Social Simulation*, 9(3):2, 2006. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/9/3/2.html`.

[65] I. Heywood, S. Conelius, and S. Carver. *An Introduction to geographical information systems*. Harlow: Pearson Education, 2012.

[66] M.I. Hoffert, K. Caldeira, G. Benford, D.R. Criswell, C. Green, H. Herzog, A.K. Jain, H.S. Kheshgi, K.S. Lackner, J.S. Lewis, H.D. Lightfoot, W. Manheimer, J.C. Mankins, M.E. Mauel, L.J. Perkins, M.E. Schlesinger, and T.M.L. Volk, T.and Wigley. Advanced technology paths to global climate stability: Energy for a greenhouse planet. *Science*, 298(5595):981–987, 2002. doi: 10.1126/science.1072357. URL `http://www.sciencemag.org/content/298/5595/981.abstract`.

[67] J. Hofierka and J. Kanuk. Assessment of photovoltaic potential in urban areas using open-source solar radiation tools. *Renewable Energy*, 34(10): 2206 – 2214, 2009. ISSN 0960-1481. doi: http://dx.doi.org/10.1016/

j.renene.2009.02.021. URL http://www.sciencedirect.com/science/article/pii/S0960148109000949.

[68] Q. Huang, D. Parker, S. Sun, and T. Filatova. Effects of agent heterogeneity in the presence of a land-market: A systematic test in an agent-based laboratory. *Computers, Environment and Urban Systems*, 41(0):188 – 203, 2013. ISSN 0198-9715. doi: http://dx.doi.org/10.1016/j.compenvurbsys.2013.06.004. URL http://www.sciencedirect.com/science/article/pii/S0198971513000616.

[69] A. Ipakchi and F. Albuyeh. Grid of the future. *Power and Energy Magazine, IEEE*, 7(2):52–62, 2009.

[70] S. Jacobsson and V. Lauber. The politics and policy of energy system transformation: Explaining the german diffusion of renewable energy technology. *Energy Policy*, 34(3):256 – 276, 2006. ISSN 0301-4215. doi: http://dx.doi.org/10.1016/j.enpol.2004.08.029. URL http://www.sciencedirect.com/science/article/pii/S0301421504002393.

[71] Adam B. Jaffe and Robert N. Stavins. Dynamic incentives of environmental regulations: The effects of alternative policy instruments on technology diffusion. *Journal of Environmental Economics and Management*, 29(3):S43–S63, 11// 1995. ISSN 0095-0696. URL http://www.sciencedirect.com/science/article/pii/S0095069685710601.

[72] Adam B. Jaffe, Richard G. Newell, and Robert N. Stavins. A tale of two market failures: Technology and environmental policy. *Ecological Eco-*

nomics, 54(23):164–174, 8/1/ 2005. ISSN 0921-8009. URL `http://www.sciencedirect.com/science/article/pii/S0921800905000303`.

[73] Kevin M. Johnston, Michael J. North, and Daniel G. Brown. *Agent Analyst: Agent Based Modeling in ArcGIS*. ESRI Press, Redlands, California, 1 edition, 2013 2013. ISBN 9781589483231.

[74] Sandeep Kapur. Technological diffusion with social learning. *The Journal of Industrial Economics*, 43(2):pp. 173–195, 1995. ISSN 00221821. URL `http://www.jstor.org/stable/2950480`.

[75] E. Katz. The social itinerary of technical change: two studies on the diffusion of innovation. *Human Organization*, 20:70–82, 1961.

[76] Peter Kaufmann, Sigrid Stagl, and Daniel W. Franks. Simulating the diffusion of organic farming practices in two new eu member states. *Ecological Economics*, 68(10):2580–2593, 8/15/ 2009. ISSN 0921-8009. URL `http://www.sciencedirect.com/science/article/pii/S0921800909001372`.

[77] H. Harry Kelejian and Ingmar R. Prucha. On the asymptotic distribution of the moran i test statistic with applications. *Journal of Econometrics*, 104(2):219–257, 9// 2001. ISSN 0304-4076. URL `http://www.sciencedirect.com/science/article/pii/S0304407601000641`.

[78] Rene Kemp and Massimiliano Volpi. The diffusion of clean technologies: a review with suggestions for future diffusion analysis. *Journal of*

Cleaner Production, 16(1, Supplement 1):14 – 21, 2008. ISSN 0959-6526. doi: http://dx.doi.org/10.1016/j.jclepro.2007.10.019. URL `http://www.sciencedirect.com/science/article/pii/S095965260700203X`. Diffusion of cleaner technologies: Modeling, case studies and policy.

[79] D. Kempe, J. Kleinberg, and . Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003. ISBN 1581137370.

[80] Lisa Klein and Gary Ford. Consumer search for information in the digital age: An empirical study of prepurchase search for automobiles. *Journal of Interactive Marketing*, 17(3):29–49, // 2003. ISSN 1520-6653. URL `http://dx.doi.org/10.1002/dir.10058`.

[81] D. Kniveton, C. Smith, and S. Wood. Agent-based model simulations of future changes in migration flows for burkina faso. *Global Environmental Change*, 21(0):S34 – S40, 2011. ISSN 0959-3780. doi: http://dx.doi.org/10.1016/j.gloenvcha.2011.09.006. URL `http://www.sciencedirect.com/science/article/pii/S0959378011001415`.

[82] Gurkan Kumbaroglu, Reinhard Madlener, and Mustafa Demirel. A real options evaluation model for the diffusion prospects of new renewable power generation technologies. *Energy Economics*, 30(4):1882 – 1908, 2008. ISSN 0140-9883. doi: http://dx.doi.org/10.1016/j.eneco.2006.10.

009. URL http://www.sciencedirect.com/science/article/pii/S0140988306001241.

[83] M. Kuperman and G. Abramson. Small world effect in an epidemiological model. *Physical Review Letters*, 86:2909–2912, Mar 2001. doi: 10.1103/PhysRevLett.86.2909. URL http://link.aps.org/doi/10.1103/PhysRevLett.86.2909.

[84] Calvin Lee Kwan. Influence of local environmental, social, economic and political variables on the spatial distribution of residential solar pv arrays across the united states. *Energy Policy*, 47(0):332–344, 8// 2012. ISSN 0301-4215. URL http://www.sciencedirect.com/science/article/pii/S0301421512003795.

[85] Carlos E. Laciana and Santiago L. Rovere. Ising-like agent-based technology diffusion model: Adoption patterns vs. seeding strategies. *Physica A: Statistical Mechanics and its Applications*, 390(6):1139–1149, 3/15/ 2011. ISSN 0378-4371. URL http://www.sciencedirect.com/science/article/pii/S0378437110009532.

[86] T. Lee, R. Yao, and P. Coker. An analysis of uk policies for domestic energy reduction using an agent based tool. *Energy Policy*, 66(0): 267 – 279, 2014. ISSN 0301-4215. doi: http://dx.doi.org/10.1016/j.enpol.2013.11.004. URL http://www.sciencedirect.com/science/article/pii/S030142151301104X.

[87] R. Lerman and S. Yitzhaki. A note on the calculation and inter-

pretation of the gini index. *Economics Letters*, 15(3-4):363 – 368, 1984. ISSN 0165-1765. doi: http://dx.doi.org/10.1016/0165-1765(84)90126-5. URL http://www.sciencedirect.com/science/article/pii/0165176584901265.

[88] J. Leskovec, L.A. Adamic, and B.A. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1(1), May 2007. ISSN 1559-1131. doi: 10.1145/1232722.1232727. URL http://doi.acm.org/10.1145/1232722.1232727.

[89] Adela Luque. An option-value approach to technology adoption in u.s. manufacturing: Evidence from microdata. *Economics of Innovation and New Technology*, 11(6):543–568, 2002. doi: 10.1080/10438590214337. URL http://www.tandfonline.com/doi/abs/10.1080/10438590214337.

[90] C. Macal and M. North. Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3):151–162, 2010. ISSN 1747-7778.

[91] Andrew Macintosh and Deb Wilkinson. Searching for public benefits in solar subsidies: A case study on the australian government's residential photovoltaic rebate program. *Energy Policy*, 39(6):3199 – 3209, 2011. ISSN 0301-4215. doi: http://dx.doi.org/10.1016/j.enpol.2011.03.007. URL http://www.sciencedirect.com/science/article/pii/S030142151100187X.

[92] T.J. Madden, P.S. Ellen, and I. Ajzen. A comparison of the theory of

planned behavior and the theory of reasoned action. *Personality and Social Psychology Bulletin*, 18(1):3–9, 1992.

[93] V. Mahajan, E. Muller, and F.M. Bass. New product diffusion models in marketing: A review and directions for research. *The Journal of Marketing*, 1:1–26, 1990.

[94] T. Mainieri, E. Barnett, T. Valdero, J. Unipan, and S. Oskamp. Green buying: The influence of environmental concern on consumer behavior. *Journal of Social Psychology*, 137(2):189–204, 1997/04/01 1997. ISSN 0022-4545. URL `http://dx.doi.org/10.1080/00224549709595430`.

[95] G.R. Maio, D.W. Bell, and V.M. Esses. Ambivalence and persuasion: The processing of messages about immigrant groups. *Journal of Experimental Social Psychology*, 32(6):513 – 536, 1996. ISSN 0022-1031. doi: http://dx.doi.org/10.1006/jesp.1996.0023. URL `http://www.sciencedirect.com/science/article/pii/S0022103196900232`.

[96] R. Margolis and J. Zuboy. Nontechnical barriers to solar energy use: Review of recent literature. National Renewable Energy Laboratory, U.S. Department of Energy, 2006. URL `www.nrel.gov/docs/fy07osti/40116.pdf`.

[97] R. Matthews, N. Gilbert, A. Roach, J. Polhill, and N. Gotts. Agent-based land-use models: a review of applications. *Landscape Ecology*, 22 (10):1447–1459, 2007. ISSN 0921-2973. doi: 10.1007/s10980-007-9135-1. URL `http://dx.doi.org/10.1007/s10980-007-9135-1`.

[98] A.J. McLane, C. Semeniuk, G.J. McDermid, and D.J. Marceau. The role of agent-based models in wildlife ecology and management. *Ecological Modelling*, 222(8):1544 – 1556, 2011. ISSN 0304-3800. doi: http://dx.doi.org/10.1016/j.ecolmodel.2011.01.020. URL `http://www.sciencedirect.com/science/article/pii/S0304380011000524`.

[99] Michael Meadows and Dave Cliff. Reexamining the relative agreement model of opinion dynamics. *Journal of Artificial Societies and Social Simulation*, 15(4):4, 2012. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/15/4/4.html`.

[100] J Stanley Metcalfe. Evolutionary economics and technology policy. *The economic journal*, 104(425):931–944, 1994.

[101] J.M. Metsaranta. Assessing factors influencing the space use of a woodland caribou rangifer tarandus caribou population using an individual-based model. *Wildlife Biology*, 14(4):478 – 488, 2008.

[102] Stanley Milgram. The small world problem. *Psychology Today*, 2(1):60 – 67, 1967.

[103] D.E. Montano and D. Kasprzyk. Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. *Health behavior and health education: Theory, research, and practice*, 4:67–95, 2008.

[104] Phillip Nelson. Information and consumer behavior. *Journal of Po-*

litical Economy, 78(2):311–329, 1970. ISSN 00223808. URL `http://www.jstor.org/stable/1830691`.

[105] Richard R Nelson and Nathan Rosenberg. *Technical innovation and national systems*, chapter 1, pages 1–18. Oxford University Press, Oxford, 1993.

[106] T. Nelson, P. Simshauser, and S. Kelley. Australian residential solar feed-in tariffs: industry stimulus or regressive form of taxation? *Economic Analysis & Policy*, 41(2):113–129, 2011.

[107] D. Noll, C. Dawes, and V. Rai. Solar community organizations and active peer effects in the adoption of residential pv. *Energy Policy*, 2014.

[108] Michael J North and Charles M Macal. *Managing business complexity: discovering strategic solutions with agent-based modeling and simulation*. Oxford University Press, 2007.

[109] J.P. Onnela, J. Saramki, J. Hyvnen, G. Szab, D. Lazer, K. Kaski, J. Kertsz, and A.L. Barabsi. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007. doi: 10.1073/pnas.0610245104. URL `http://www.pnas.org/content/104/18/7332.abstract`.

[110] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: bringing order to the web. Technical Report 1999-66, Stanford

InfoLab, November 1999. URL http://ilpubs.stanford.edu:8090/422/.

[111] H. Parry, C. Topping, M. Kennedy, N. Boatman, and A. Murray. A bayesian sensitivity analysis applied to an agent-based model of bird population response to landscape change. *Environmental Modelling & Software*, 45(0):104 – 115, 2013. ISSN 1364-8152. doi: http://dx.doi.org/10.1016/j.envsoft.2012.08.006. URL http://www.sciencedirect.com/science/article/pii/S1364815212002265. ¡ce:title¿Thematic Issue on Spatial Agent-Based Models for Socio-Ecological Systems¡/ce:title¿.

[112] J.M. Pearce. Photovoltaics - a path to sustainable futures. *Futures*, 34(7):663–674, 2002. ISSN 0016-3287. doi: http://dx.doi.org/10.1016/S0016-3287(02)00008-3. URL http://www.sciencedirect.com/science/article/pii/S0016328702000083.

[113] Renana Peres, Eitan Muller, and Vijay Mahajan. Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing*, 27(2): 91 – 106, 2010. ISSN 0167-8116. doi: http://dx.doi.org/10.1016/j.ijresmar.2009.12.012. URL http://www.sciencedirect.com/science/article/pii/S0167811610000236.

[114] R.G. Pontius. Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing*, 66(8):1011–1016, 2000.

[115] David Popp, Ivan Hascic, and Neelakshi Medhi. Technology and the diffusion of renewable energy. *Energy Economics*, 33(4):648–662, 7// 2011. ISSN 0140-9883. URL http://www.sciencedirect.com/science/article/pii/S0140988310001283.

[116] G Premkumar, Keshavamurthy Ramamurthy, and Sree Nilakanta. Implementation of electronic data interchange: an innovation diffusion perspective. *Journal of Management Information Systems*, 11(2):157–186, 1994.

[117] Hazhir Rahmandad and John Sterman. Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science*, 54(5):998–1014, May 1, 2008 2008. URL http://mansci.journal.informs.org/content/54/5/998.abstract.

[118] V. Rai and S.A. Robinson. Effective information channels for reducing costs of environmentally- friendly technologies: evidence from residential pv markets. *Environmental Research Letters*, 8(1):014044, 2013. ISSN 1748-9326. URL http://stacks.iop.org/1748-9326/8/i=1/a=014044.

[119] V. Rai and B. Sigrin. Diffusion of environmentally-friendly technologies: Buy vs. lease decisions in residential pv markets. *Environmental Research Letters*, 8(1):1–8, 2012.

[120] Varun Rai and Kristine McAndrews. Decision-making and behavior

change in residential adopters of solar pv. In *World Renewable Energy Forum*, Denver, CO, May 2012.

[121] S. Railsback, S. Lytinen, and S. Jackson. Agent-based simulation platforms: Review and development recommendations. *Simulation*, 82(9): 609–623, 2006. doi: 10.1177/0037549706073695. URL http://sim.sagepub.com/content/82/9/609.abstract.

[122] W. Rand and R.T. Rust. Agent-based modeling in marketing: Guidelines for rigor. *International Journal of Research in Marketing*, 28(3): 181 – 193, 2011. ISSN 0167-8116. doi: http://dx.doi.org/10.1016/j.ijresmar.2011.04.002. URL http://www.sciencedirect.com/science/article/pii/S0167811611000504.

[123] T. Rashidi and A. Mohammadian. A competing hazard model of household vehicle transaction behavior with discrete time intervals and unobserved heterogeneity. *Transportation Letters*, 3(3):219–229, 2011. doi: 10.3328/TL.2011.03.03.219-229. URL http://www.maneyonline.com/doi/abs/10.3328/TL.2011.03.03.219-229.

[124] J. Reiss. A plea for (good) simulations: nudging economics toward an experimental science. *Simulation & gaming*, 42(2):243–264, 2011.

[125] J. Richetin, A. Sengupta, M. Perugini, I. Adjali, R. Hurling, D. Greetham, and M. Spence. A micro-level simulation for the prediction of intention and behavior. *Cognitive Systems Research*, 11(2): 181 – 193, 2010. ISSN 1389-0417. doi: http://dx.doi.org/10.1016/j.

cogsys.2009.08.001. URL http://www.sciencedirect.com/science/article/pii/S1389041709000400.

[126] S. Robinson, M. Stringer, V. Rai, and A. Tondon. Gis-integrated agent-based modeling of residential solar pv diffusion. In *In: 32nd USAEE/IAEE North American Conference Proceedings*, July 2013.

[127] E.M. Rogers. *Diffusion of Innovations*. Free Press, New York, NY, 2003.

[128] Michael Rothschild. Models of market organization with imperfect information: a survey. *The Journal of Political Economy*, pages 1283–1308, 1973.

[129] S. Schnettler. A structured overview of 50 years of small-world research. *Social Networks*, 31(3):165 – 178, 7// 2009. ISSN 0378-8733. URL http://www.sciencedirect.com/science/article/pii/S0378873308000695.

[130] N. Schwarz and A. Ernst. Agent-based modeling of the diffusion of environmental innovations: An empirical approach. *Technological Forecasting and Social Change*, 76(4):497 – 511, 2009. ISSN 0040-1625. doi: 10.1016/j.techfore.2008.03.024. URL http://www.sciencedirect.com/science/article/pii/S0040162508000875. ¡ce:title¿Evolutionary Methodologies for Analyzing Environmental Innovations and the Implications for Environmental Policy¡/ce:title¿.

[131] M. Schwoon. Simulating the adoption of fuel cell vehicles. *Journal of Evolutionary Economics*, 16(4):435 – 472, 2006. ISSN 0936-9937.

[132] J. Scott and P.J. Carrington. *The SAGE Handbook of Social Network Analysis*. SAGE publications, 2011.

[133] R. Shrestha and C. Gopalakrishnan. Adoption and diffusion of drip irrigation technology: An econometric analysis. *Economic Development and Cultural Change*, 41(2):407–418, 1993. ISSN 00130079. URL `http://www.jstor.org/stable/1154429`.

[134] A. Smajgl and O. Barreteau. *Empirical Agent-Based Modeling-Challenges and Solutions*. Springer, 2014.

[135] A. Smajgl and E. Bohensky. Behaviour and space in agent-based modelling: Poverty patterns in east kalimantan, indonesia. *Environmental Modelling & Software*, 45(0):8 – 14, 2013. ISSN 1364-8152. doi: http://dx.doi.org/10.1016/j.envsoft.2011.10.014. URL `http://www.sciencedirect.com/science/article/pii/S1364815211002258`. ¡ce:title¿Thematic Issue on Spatial Agent-Based Models for Socio-Ecological Systems¡/ce:title¿.

[136] A. Smajgl, D. Brown, D. Valbuena, and M. Huigen. Empirical characterisation of agent behaviours in socio-ecological systems. *Environmental Modelling & Software*, 26(7):837 – 844, 2011. ISSN 1364-8152. doi: http://dx.doi.org/10.1016/j.envsoft.2011.02.011. URL `http://www.sciencedirect.com/science/article/pii/S1364815211000557`.

[137] Eliot R Smith and Frederica R Conrey. Agent-based modeling: A new approach for theory building in social psychology. *Personality and social psychology review*, 11(1):87–104, 2007.

[138] B.M. Sopha, C.A. Klockner, and E.G. Hertwich. Adoption and diffusion of heating systems in norway: Coupling agent-based modeling with empirical research. *Environmental Innovation and Societal Transitions*, 8(0):42 – 61, 2013. ISSN 2210-4224. doi: http://dx.doi.org/10.1016/j.eist.2013.06.001. URL `http://www.sciencedirect.com/science/article/pii/S2210422413000427`.

[139] George J. Stigler. The economics of information. *Journal of Political Economy*, 69(3):213–225, 1961. ISSN 00223808. URL `http://www.jstor.org/stable/1829263`.

[140] Paul Stoneman and NJ Ireland. The role of supply factors in the diffusion of new process technology. *The Economic Journal*, 93:66–78, 1983.

[141] David Strang and Sarah A Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual review of sociology*, pages 265–290, 1998.

[142] S. Thiriot and J.D. Kant. Using associative networks to represent adopters' beliefs in a multiagent model of innovation diffusion. *Advances in Complex Systems*, 11(02):261–272, 2008. doi: 10.1142/S0219525908001611. URL `http://www.worldscientific.com/doi/abs/10.1142/S0219525908001611`.

[143] M. Tran. Agent-behaviour and network influence on energy innovation diffusion. *Communications in Nonlinear Science and Numerical Simulation*, 2012. ISSN 1007-5704.

[144] Peter Turchin, Thomas E. Currie, Edward A. L. Turner, and Sergey Gavrilets. War, space, and the evolution of old world complex societies. *Proceedings of the National Academy of Sciences*, 110:16384 – 16389, 2013. doi: 10.1073/pnas.1308825110. URL `http://www.pnas.org/content/early/2013/09/20/1308825110.abstract`.

[145] Thomas W. Valente. Social network thresholds in the diffusion of innovations. *Social Networks*, 18(1):69 – 89, 1996. ISSN 0378-8733. doi: http://dx.doi.org/10.1016/0378-8733(95)00256-1. URL `http://www.sciencedirect.com/science/article/pii/0378873395002561`.

[146] H. Van Dyke Parunak, Robert Savit, and RickL. Riolo. Agent-based modeling vs. equation-based modeling: A case study and user's guide. In JaimeSimo Sichman, Rosaria Conte, and Nigel Gilbert, editors, *Multi-Agent Systems and Agent-Based Simulation*, volume 1534 of *Lecture Notes in Computer Science*, pages 10–25. Springer Berlin Heidelberg, 1998. ISBN 978-3-540-65476-6. doi: 10.1007/10692956_2. URL `http://dx.doi.org/10.1007/10692956_2`.

[147] R. Van Haaren and V. Fthenakis. Gis-based wind farm site selection using spatial multi-criteria analysis (smca): Evaluating the case for new york state. *Renewable and Sustainable Energy Reviews*, 15(7):

3332 – 3340, 2011. ISSN 1364-0321. doi: http://dx.doi.org/10.1016/j.rser.2011.04.010. URL `http://www.sciencedirect.com/science/article/pii/S136403211100147X`.

[148] H. Visser and T. de Nijs. The map comparison kit. *Environmental Modelling & Software*, 21(3):346 – 358, 2006. ISSN 1364-8152. doi: http://dx.doi.org/10.1016/j.envsoft.2004.11.013. URL `http://www.sciencedirect.com/science/article/pii/S1364815204003019`.

[149] Oscar van Vliet, Bert de Vries, Andre Faaij, Wim Turkenburg, and Wander Jager. Multi-agent simulation of adoption of alternative fuels. *Transportation Research Part D: Transport and Environment*, 15(6):326–342, 8// 2010. ISSN 1361-9209. URL `http://www.sciencedirect.com/science/article/pii/S1361920910000374`.

[150] D. Watts and S. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

[151] Stephen R. Wealands, Rodger B. Grayson, and Jeffrey P. Walker. Quantitative comparison of spatial fields for hydrological model assessment–some promising approaches. *Advances in Water Resources*, 28(1):15–32, 2005. ISSN 0309-1708. doi: http://dx.doi.org/10.1016/j.advwatres.2004.10.001. URL `http://www.sciencedirect.com/science/article/pii/S0309170804001575`.

[152] U. Wilensky. *NetLogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1999.

[153] H. Peyton Young. Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *The American Economic Review*, 99(5):pp. 1899–1924, 2009. ISSN 00028282. URL http://www.jstor.org/stable/25592541.

[154] T. Zhang. Agent-based simulation of consumer purchase decision-making and the decoy effect. *Journal of Business Research*, 60(8): 912 – 922, 2007. ISSN 0148-2963. doi: 10.1016/j.jbusres.2007.02.006. URL http://www.sciencedirect.com/science/article/pii/S0148296307000483. ¡ce:title¿Complexities in Markets Special Issue¡/ce:title¿.

[155] T. Zhang and W. J. Nuttal. An agent-based simulation of smart metering technology adoption. *International Journal of Agent Technologies and Systems*, 4:17–38, 2012.

[156] B. P. Ziegler. *Theory of Modeling and Simulation*. Wiley, New York, 1976.