Agent-Based Modeling of Energy Technology Adoption: Empirical Integration of Social, Behavioral, Economic, and Environmental Factors

Varun Rai*[1,2] and Scott A. Robinson[1,3]

[1]LBJ School of Public Affairs, The University of Texas at Austin
[2]Department of Mechanical Engineering, The University of Texas at Austin
[3]Jackson School of Geosciences, The University of Texas at Austin

February 26, 2015

Abstract

Agent-based modeling (ABM) techniques for studying human-technical systems face two important challenges. First, agent behavioral rules are often *ad hoc*, making it difficult to assess the implications of these models within the larger theoretical context. Second, the lack of relevant empirical data precludes many models from being appropriately initialized and validated, limiting the value of such models for exploring emergent properties or for policy evaluation. To address these issues, in this paper we present a theoretically-based and empirically-driven agent-based model of technology adoption, with an application to residential solar photovoltaic (PV). Using household-level resolution for demographic, attitudinal, social network, and environmental variables, the integrated ABM framework we develop is applied to real-world data covering 2004-2013 for a residential solar PV program at the city scale. Two applications of the model focusing on rebate program design are also presented.

Keywords: Agent-based modeling; Solar photovoltaic (PV); Complex systems; Technology adoption; Social networks; Bounded rationality

1 Introduction

Development of methods that better represent the bounded rationality of economic agents [39], largely arising due to heterogeneous information sets and heuristic decision-making [22], is important for better understanding of the emergent phenomena that permeate economic systems [90, 91]. In this vein, recent years have seen a spurt in the use of agent-based modeling

*raivarun@utexas.edu. 2315 Red River St., LBJ School of Public Affairs, The University of Texas at Austin, Texas, 78712.

(ABM) in a range of economic and human-technical systems, including transportation [108], land use [33, 88], market structure [50, 58], transaction costs [115], strategic interactions in climate policy [14, 38], and technology adoption [94], especially that of environmentally-friendly technologies [17, 47, 60, 61, 70, 93, 101, 106, 117]. ABM is attractive to researchers interested in studying the evolution of complex human-technical systems because of the flexibility afforded by ABM to describe in great detail the behavioral as well as structural (policy; prices; infrastructure) aspects of the system. However, ABM techniques for studying human-technical systems face two important challenges [30, 113]. First, agent behavioral rules in agent-based models are often *ad hoc* – they do not necessarily build upon systematic theories of behavior, thereby making it difficult to assess the implications of these models within the larger theoretical context [30, 35]. Second, the lack of relevant empirical data precludes many models from being appropriately initialized and validated against real-world data [50]; this limits the value of such models for exploring emergent properties or for policy evaluation. Thus, careful development of agent behavioral model and of rich datasets and methods to enable robust initialization and validation of agent-based models is important.

1.1 Objectives

Our objective in this paper is to present an agent-based model of technology adoption that systematically tries to address the challenges identified above regarding the theoretical and empirical components of ABM. Using a uniquely rich and comprehensive dataset covering 2004-2013, we build an agent-based model of the adoption of residential solar photovoltaic (PV) systems in the city of Austin (Texas, USA), which has a population of approximately 900,000. In addition to an empirically driven agent-interaction model and a theoretically-driven behavioral model, we also account in great detail for the physical environment (irradiation, tree cover, home size) and economic features (prices, subsidies, wealth) that impact agent behavior, again using empirical data. We present a detailed step-by-step construction of the different components of the agent-based model, the process of initializing the model, and setting up of the model parameters and variables in accordance with the theoretical and empirical underpinnings of the system. We also present methods for temporal and spatial validation of the model along with the fitting and validation results. The emphasis is to provide a high level of detail in the construction of the behavioral model, data integration, and initialization procedures. These details, which are critical to making the model reproducible, are often neglected [45]. Because agent-based models intended for policy evaluation, predictive modeling, or the study of emergent phenomena must go through a rigorous model set-up and empirical grounding, we hope that this paper will help facilitate the development of agent-based models for energy technology adoption in a more empirically-grounded fashion.

1.2 Applications

A key motivation in developing this framework is to allow for a range of policy simulations that could inform decision-making of policymakers and utility planners. To illustrate this potential, in Section 4.2 we present two applications of the ABM framework developed here. These policy scenarios are based on the following more general questions of central importance to the designers of solar programs, but the framework is generalizable across a suite of technologies.[1]

Subsidy Program Design

Low-Income Solar Programs: Adopters of PV tend to be much wealthier than average [82]. This finding has raised equity concerns in relation to publicly-funded rebates. Programs like California's Single-Family Affordable Solar Housing Program were created to address these concerns, but high cost and long time-frames associated with solar PV adoption limit the ability of program designs to experiment with different rebate offerings. Using the framework developed here a range of *targeted rebate* scenarios could be explored through ABM simulation experiments.

Rebate Levels and Adoption: Recent empirical findings on optimal subsidy design suggest that when peer effects and learning-by-doing effects are strong rebates should be front-loaded in order to maximize adoption – larger rebates early on in the subsidy program and declining over time are found to be more cost effective [29, 105]. These aggregate findings could be validated in full-scale ABM simulations including two- or multi- tiered rebates and by varying the rate at which the tiered rebates change over the lifetime of a solar program. The key outcome of interest is the elasticity of PV adoption: (i) what is the impact of changes in rebate level on PV adoption at the population-scale? and (ii) how does this impact change with the underlying installation base?

1.3 Main Contributions

The main contributions of this paper are: (i) development of a theoretically and empirically grounded integrated model for consumer technology adoption, applied to residential solar PV, (ii) highly granular description of the system, including behavioral, social, and physical-economic environmental aspects at the household level, (iii) development of new techniques to achieve a population-wide, household-level empirical initialization, (iv) development and application of multiple (temporal, spatial, and demographic) external validation metrics, and (v) application of the developed model for two ABM simulation experiments to explore solar program design.

[1]Our framework may be applied not only to solar PV but also to a range of other consumer technologies. We provide the applications for the design of solar programs because the empirical components of our model are trained on granular data from a solar program. Similar data on other technologies would enable studying the adoption of those technologies as well. Furthermore, although not considered in this paper, other simulation experiments could include exploration of different information seeding strategies and location-based rebate targeting.

2 Background and Related Literature

Advances in computing power combined with the increasing availability of granular data have enabled researchers to apply ABM for analyzing a diverse set of problems [4, 69]. A particular area of growth in ABM applications has been to model consumer technology adoption, a problem for which standard methods include conjoint analysis [31, 44], Bass diffusion models [52, 53, 95], and dynamic discrete choice (DDC) models [9]. Modeling of consumer energy technology adoption is particularly challenging because the nominal economics (price) of the technology is only one determinant of consumers' likelihood to adopt. Other behavioral and social phenomenon such as decision heuristics, anchoring, path-dependence (past experiences), risk aversion, trust-based information networks, and social norms are also quite important in understanding energy-related consumer decision-making [27, 43, 57, 68, 99, 112]. DDC models are among the most sophisticated approaches for analyzing consumer choice [71]. Unlike conventional conjoint analysis, DDC models have a time component (multi-period), allowing to factor intertemporal tradeoffs. Unlike Bass diffusion models, the unit of analysis in DDC models is the individual, thereby allowing the direct study of individual decision-making processes on system outcomes. However, these predominant methods for modeling consumer technology adoption often rely heavily on assumptions of utility maximizing actors who have rational expectations about the future technological trajectory. Furthermore, there are several other key challenges associated with the representation of important behavioral, social, and spatial phenomena in conventional models of energy technology adoption (see the review in [57]).

While the potential of ABM to address the weaknesses of conventional diffusion models is quite promising, it is important that ABM development for the study of human-technical systems follow fundamentally sound principles [30, 46, 78, 86, 96]. In particular, agents' decision rules [30], the empirical basis of the system description [12, 79, 97, 98], and model validation [34, 50, 111] demand rigorous treatment. This is especially important if policy evaluation or predictive modeling is the main objective. Though not always followed in practice, the need for empirical basis and validation in ABMs has been recognized before [30, 36, 46, 86]. By grounding the agent states, decision rules, and environmental variables in empirical patterns, ABMs could gain descriptive [32], explanatory [30], and predictive power [85].

Mindful of the opportunities and pitfalls of ABM, in this paper we present a theoretically-driven and empirically-grounded ABM of residential solar adoption. Our choice to study solar adoption using ABM is motivated primarily by two factors. First, over the last decade solar has emerged as a serious electricity supply option, and has been the fastest growing energy technology globally [37, 102, 103]. Solar's spread also has important broader implications. For example, while the penetration of solar is still quite low in most of the United States (and the rest of the world), solar has already upended conventional expectations and has been at the center of discussions on revisiting the conventional electric utility model in the U.S. [11]. Second, from a theoretical viewpoint, acquiring a solar system is a complex decision requiring significant time and monetary resources from the consumer [83]. Further, the decision to adopt solar is characterized by "non-price" interactions between consumers with limited (and different) information sets, which in turn are fed via different (individually specific) social networks. For example, some recent research provides strong evidence about the importance of local peer-networks in driving the rate of adoption of residential solar [13, 77, 83]. As such, studying the solar adoption process provides a unique window into exploring the structure and role of social interactions embedded within such complex decisions. ABM is a naturally suitable method to represent these complexities of the solar adoption decision process. The rest of this paper is a detailed presentation of the data and methodology used to set up, initialize, validate the solar ABM, and discuss its applications.

3 Data and Methodology

Our modeling goal is to build a *household-level agent-based model* that is able to generate the empirically observed temporal and spatial patterns of the adoption of residential solar at the city scale. Accordingly, the solar adoption ABM presented here uses the 173,466 actual single-family residential households in Austin, Texas (accurate as of mid 2013) as the primary agents. The time period of the simulation is from 2004 to June 2013, during which the adoption level grew from only $N = 20$ to 2,738. Note that since the adoption level (1.58% of the relevant population) at the end of the model period is relatively low, it makes the modeling task especially challenging. In this section we present the step-by-step process of the solar ABM: (i) model set up and how the different model components interact with each other, and (ii) the integration of granular data from surveys, utility rebate programs, and household-level publicly available data to represent the environment and to initialize agent states. The simulations were run using the Stampede Supercomputer at the Texas Advanced Computing Center (TACC).

3.1 Data

In order to accurately represent market conditions, agent-based models need to incorporate empirical data for initialization and validation. In this section we provide details for each of the data-streams in the model and then elaborate in the later sections upon the way data are used in the model. As explained below, the resulting household-level database is a combination of granular socio-economic demographics and environmental data for *every household* in the study area as well as detailed time-series rebate program and survey data for the solar adopters. This combination of carefully overlaid data-streams is critical in enabling us to model the economic, attitudinal, and social network attributes.

3.1.1 Austin Solar Rebate Program Data

Program tracking data on PV adopters was collected by the electric utility (Austin Energy) as a part of the implementation of Austin's solar rebate program. This data included information about the installation date, system size, all technical details of the system, rebate amount, system cost, and the location of the installation. Installation locations were

Table 1: Data streams incorporated into the ABM for initialization, fitting, and validation. Scale refers to the population of agents used in the model, which covers the 173,466 residential households in the city of Austin.

Data Sources					
Source	**Content**	**Scope**	**Scale**	**Resolution**	**Time frame**
Austin Energy	System details, location, installation date, etc	PV Adopters	Population	Household	2004-2013
UT Austin	Survey responses regarding installation decision	PV Adopters	Sample, 22.5%	Household	2011-2014
TCAD	Home value, parcel size, land use code, etc	Land parcels	Population	Household	2013
CAPCOG	LIDAR images	City of Austin visible above ground	Population	6in	2013
USGS	National Elevation Dataset	City of Austin elevation ASL	Population	3m	2013

geocoded to street locations in a GIS with a 97.9% match rate. Geocoding the matched locations allowed geographic distributions of PV system characteristics and survey responses (see below) to be overlaid with other socio-economic demographics and environmental layers.

3.1.2 Solar Adopter Survey

Longitudinal survey data on PV adopters in Austin were collected in three main waves from 2011-2014. The 616 responses from PV adopters in Austin constituted a 22.5% response rate. The survey consisted of detailed questions regarding the behavioral, financial, and social components of the adoption decision. Relevant questions are included in Appendix A. Further, respondents were given the option to provide personal identifiers, allowing responses to be joined to detailed solar rebate program data (Section 3.1.1). 82% of respondents opted to provide this additional information.

3.1.3 Appraisal District Data

Publicly available data from the Travis County Appraisal District (TCAD) were used to generate household parcel polygon shapefiles, which were joined to the TCAD database containing appraised and market home value, parcel area, land-use codes, construction date, and major home improvements. Land-use codes were used to filter the data down to single-family residential parcels. A spatial join was used to match geocoded PV adopter locations to residential parcels.

3.1.4 Household Footprint, Tree-cover, and Terrain

Publicly available light detection and ranging (LIDAR) data from the City of Austin were used to generate household footprint and tree-cover layers. One LIDAR raster was used, as time-series LIDAR data were not available. The household footprints were approximated using the roof area shape. Tree-cover was assessed using the IR band. These layers were overlaid and joined to the residential households. Capital Area Council of Governments (CAPCOG) data were also used to define utility service area boundaries, zip codes, roadways, and bodies of water. Terrain and elevation data from the U.S. Geological Survey (USGS) National Elevation Dataset at the 3 meter level were used to create a digital elevation model (DEM) of the study area. DEM derivatives such as slope, curvature, aspect, and hillshade were calculated and used to calculate solar irradiance in Watt-hours per square foot for the entire study area.

3.2 Model Overview and Design

The formulation of our behavioral model is motivated by the Theory of Planned Behavior (TPB) – a widely applied behavioral model in psychology [2, 5, 41, 64].[2] Variations of the TPB framework have been applied in a number of agent-based models describing theoretical markets [116], human migration [59], dietary choice [87], and technology diffusion [55, 93, 98, 117]. Figure 1 shows the various components of the solar ABM developed here. These components were combined in the integrated model in the R programming language, with supporting methods written in Python. Each of these components is described in detail later in this section. We begin by providing an overview and design concepts of the model components to illustrate the mechanics of the ABM.

In the model two key elements determine the decision of agents to adopt or not adopt solar: an attitudinal component ("attitude") and a control component ("control"). As explained later, a social network model is embedded within the Attitudinal module. As the model cycles forward in time, both the attitude and control attributes of all agents evolve based on interactions with other agents (in respective social networks) and/or feedback from the environment. The *decision criteria* is that both an agent's attitude and control attributes must be above certain respective thresholds before she adopts solar. We use a *global attitude threshold* [104] to determine if an agent has a strong enough positive attitude toward the

[2]TPB states that human behavior is the result of the *intention* to perform the behavior. In turn, the intention itself is driven by the individual's *attitude* toward the behavior, *subjective norms*, i.e., perceptions about social expectations and pressure, and *perceived behavioral control* (PBC), i.e., the individual's perception of her ability to actually perform the behavior [3]. Thus, "[a]s a general rule, the more favorable the attitude and subjective norm, and the greater the perceived control, the stronger should be the person's intention to perform the behavior in question" [3]. The three components of intention can be modeled as a function of attitudinal, social, and demographic variables [66, 75].

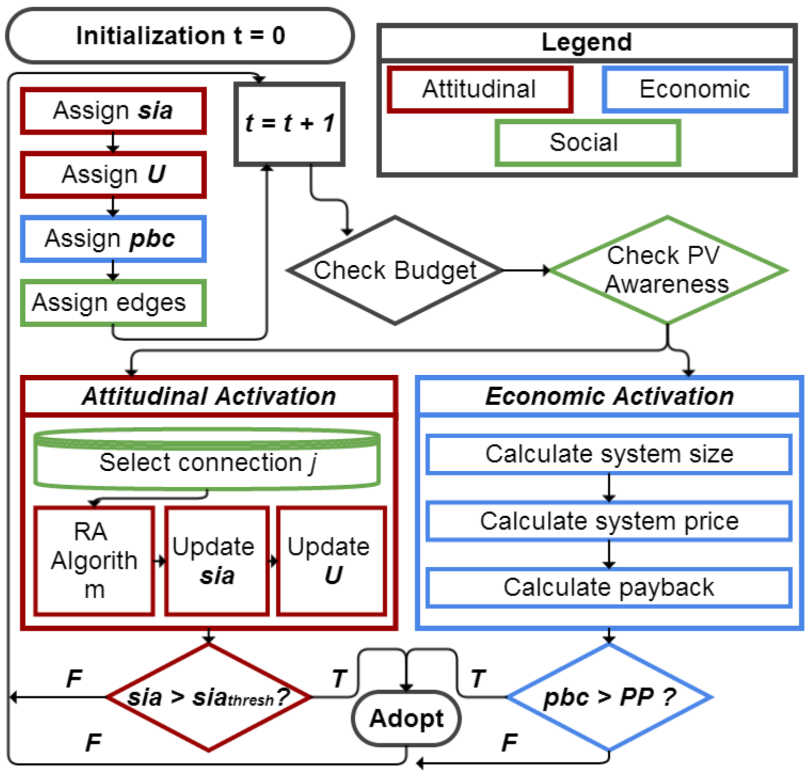


Figure 1: Flowchart describing the solar ABM structure, emphasizing the agent decision process to install solar. Agent states are initialized using a population-wide empirical process using data through Q4 2007. Thus, the first model cycle occurs in Q1 2008. Agents' attitudes (sia) and uncertainty regarding those attitudes (U) are modified through interactions with other agents in their small-world social network through the Relative Agreement algorithm, and compared to a global threshold (sia^{thresh}). Individual control beliefs (pbc) regarding ability to afford solar are compared to current payback periods (PP), which are influenced by house location, electricity prices and available incentives. Adoption occurs only when *both* the attitudinal and the economic criteria are met. The "Check PV Awareness" and "Check PV Awareness" steps exist in the framework, but do not impact the model outcomes.[3]

technology to potentially adopt it. Section 3.3.3 presents the detailed methodology for the attitudinal sub-model. To determine whether an agent can potentially afford to adopt the technology, we use an *individual control threshold*. Details of the control sub-model are presented in Section 3.3.1. Following the steps, if the agent has a sufficiently strong attitude *and* control, then the agent will adopt solar. Note that this formulation specifies the attitude and control components as being able to evolve independently. This is consistent with the underlying construction of TPB. An examination of the data for our empirical case (solar adoption in Austin, TX) supports this formulation. Further, a recent survey of non-adopters in Texas confirms that attitude and control variables regarding solar exhibit significant independence [81].

In its standard form TPB is formulated as a static model of behavior: *at a specific time*, TPB maps measures of attitude, subjective norms, and perceived behavioral control (PBC) onto intention, and intention onto actual behavior. TPB does not specify how these variables *evolve* over time allowing intention to change. Since in reality these measures are not static – for example, norms and attitudes evolve over time through social interactions [7, 20] – incorporating TPB in ABM requires complementary means of evolving agent variables. In our formulation, the dynamic aspect in the control component of agent behavior comes from the changing economics of solar. Section 3.3.1 discusses the control module in further details. Furthermore, as described in Section 3.3.3, we use the Relative Agreement (RA) algorithm to model the process through which agent attitudes and uncertainties around those attitudes evolve through agent-agent interactions [25, 48, 73].

3.3 Initialization

Agent states should reflect the conditions at model initiation ($t = 0$) in order to anchor the model in the empirical time-series. We match these attributes to the relevant data in the city of Austin in Q4 2007 through population-wide household-level empirical initialization. Each time step corresponds to one quarter year. Thus, in terms of time steps, Q4 2007 is $t = 0$ and Q1 2008 is $t = 1$, the first step in the model cycle. Initial agent state distributions of the fully empirical initialization process for agents economic, network, and attitudinal attributes are discussed next.

3.3.1 The Control Module: PV Economics and Payback

The control module is modeled after the perceived behavioral control (*pbc*) construct from TPB. Note that PBC in TPB is a more general concept than just a measure of economic

[3]While the model was designed to operate using a utility budget constraint if relevant ("Check Budget"), all the scenarios described in this paper are unconstrained. Furthermore, the "Check PV Awareness" step checks at each time step for each agent if they are "aware", which is defined as at least one other agent in the social network being a PV adopter. This step was included for consistency with the literature [8], but it has no significant impact on the model for two reasons: (i) within the first few time steps a vast majority of the agents become "aware" and (ii) those agents that are "unaware" are also highly likely to be either below the attitudinal or the control threshold. That is, compared to "awareness", *sia* and *pbc* are much tougher constraints on adoption, thus making the awareness check redundant.

control over a decision. For solar, however, the perception of affordability (or lack thereof) is often cited as the most important barrier to adoption [81, 84]. Thus, for our purposes focusing on the economic component of behavioral control is justified. We use the *pbc* variable as a measure of an agent's perception of her ability (control) to perform a behavior, in the face of "the presence of factors that may facilitate or impede performance of the behavior" [3]. Simple payback was the most commonly used financial metric by solar adopters surveyed in the study area [82, 84]. Accordingly, we compute pbc_i for each agent i as the minimum tolerable payback period of investing in a solar system. The interpretation in the context of TPB is that an agent perceives *full control* over adoption *if* the payback is lower than the agent's *pbc*. More concretely, an agent i compares her pbc_i with the empirical payback at the current time period PP_{it}. Assuming that she is already above the attitudinal threshold (see Section 3.3.3), adoption happens only if the payback is lower than pbc_i, that is:

$$PP_{it} < pbc_i \ . \tag{1}$$

As explained below, note that pbc_i is computed only once for each agent and has that same value throughout the simulation period; thus, we assume the *pbc* to be an intrinsic, time invariant attribute of an agent over the simulation period. In contrast, as discussed next, payback is a dynamic quantity and changes for each agent over time.

As shown in Equation 2, payback is calculated as a function of the value of the electricity produced by the solar system (e), the per unit price p of the solar system (in \$/Watt), utility rebates (R), and the federal investment tax credit (ITC) for each time period t, and the annual system electricity generation G (in kWh/kW, calculated based on site-specific irradiance):

$$PP_{it} = (p_t - R_t - (p_t - R_t) \times ITC_t)/(G_i \times e_t) \ . \tag{2}$$

In this formulation, PP is only affected by changes in prices and rebates. PP is computed independently of any physical constraint owing to tree-cover, which, as described below, is accounted for through *pbc*. Except p_t all other variables in Equation 2 are known exactly for each agent. p_t was modeled using non-parametric local polynomial regression (LOESS). In the LOESS model, price, the dependent variable, is a function of time only. The price data were compiled from Austin Energy solar rebate program tracking data. Observed prices range from nearly \$10/Watt in 2008 to just over \$2/Watt in 2013. 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, and residuals were approximately normal. The LOESS model outcome is presented in the *Supplementary Information* (*SI*).

We model agents' *pbc* as a function of their financial resources and the relevant physical features of their house. We take the home value (W) as a proxy for the financial resources available to the agent. The relevant house-feature quantities are size of the house (s), tree-cover (T), and irradiance received (I). Tree-cover and the amount of sunlight received (including any hill shade) may be expected to impact the perception of the financial viability of installing solar. This is supported by the fact that modeling *pbc* as a function of W alone leads to over-prediction of adoption among wealthy households, particularly in hilly areas.

Because *pbc* is taken as a single index value, the assumption is that given a sufficiently financially attractive payback period, the agent will take measures to overcome tree-cover related physical constraints.[4] Examining the data we find that this indeed is empirically the case as demonstrated in wealthy adopter households with high tree-cover. As shown in Equation 3, we model *pbc* as a linear sum, wherein greater financial resources and amount of sunlight received increase the agent's *pbc*, while tree-cover share over the roof decreases it:

$$pbc_i = \alpha_0 + \alpha_1 \left(I_i + W_i^* - \left(\frac{T}{s}\right)_i^* \right) , \tag{3}$$

where W^* and $(T/s)^*$ denote the weighted W and T/s (the tree-cover ratio), respectively. The weighting is necessary in order to account for differences in the scales of each of the components: W and T/s were made comparable to irradiance by assigning weights such that the medians for W^* and $(T/s)^*$ were equal to the median of I. As discussed in Section 3.6, α_0 and α_1 are parameters in the model that are fit to minimize the error between the empirical number of cumulative adoptions each quarter and the predicted number. The optimal values obtained from the fitting process are $\alpha_0 = -60.61$ and $\alpha_1 = 2.46$.

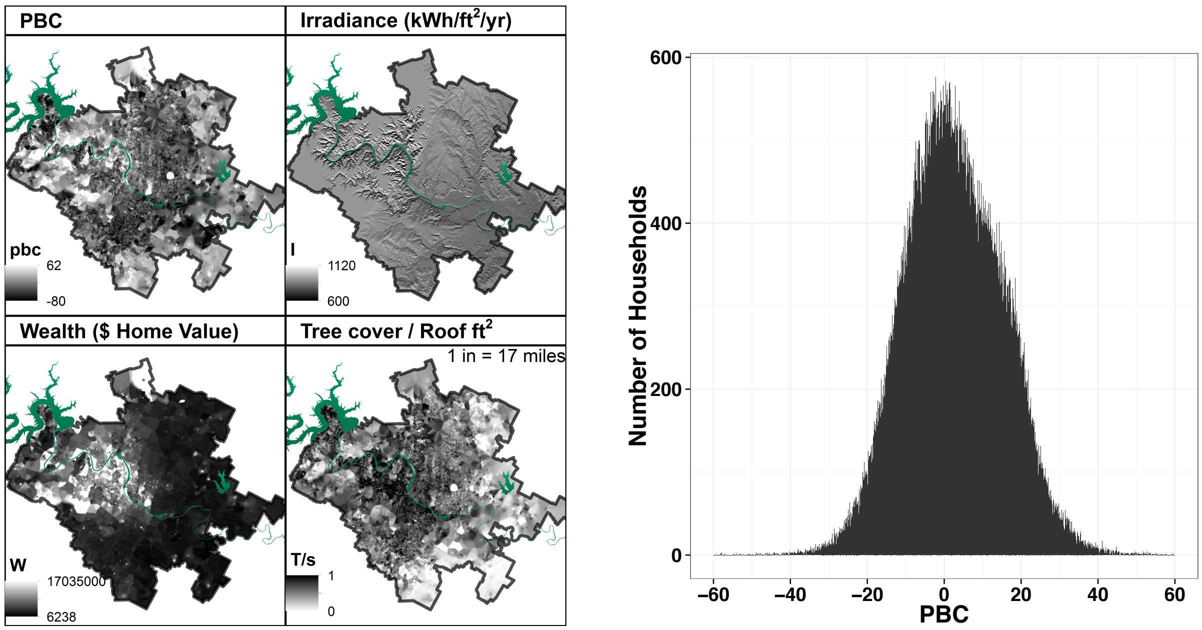
3.3.2 The Control Metric, *pbc*

By incorporating multiple relevant components, *pbc* more fully captures the complexity of solar economics. The decomposition of *pbc* components and the resultant *pbc* distribution are shown in Figure 2. As can be seen, the *pbc* index has a distinct spatial distribution (Figure 2a, top left map) from any one of its components (Figure 2a, irradiance, home value, and tree-cover maps). Note that through the fitting process described above a portion of the agent population gets assigned negative *pbc* values (Figure 2b). This reflects households that, given their financial and house attributes, would not adopt solar regardless of the payback. For instance, this could be because of need for large roofing improvements or extensive tree trimming, which would result in large costs in addition to that of the solar system alone. Looked at differently, values of *pbc* below 0 suggest that without making some of the variables dynamic that we hold constant in the model (home value, roof size, tree cover, irradiance) over the simulation time frame[5], these households will not adopt PV regardless of economics. As such, only households with positive *pbc* can be considered to have "technical" solar PV potential.

In order to measure the accuracy of the *pbc* metric, we compare the *pbc* value (Equation 3) for each PV-adopter household to that same household's *realized* payback period. The realized payback is calculated using Equation 2, but using the actual price and timing information relevant for the specific agent (this data is available from the solar rebate program dataset). Thus, the realized payback is directly observed for all PV adopters. Recall that the economic threshold rule is that an agent is able to install solar only if, at a given

[4]Unless very onerous, tree-trimming to clear tree shade for installing solar is not expected to change the economics of solar installation.

[5]These variables are held constant in the model due to lack of good historical data on which to base a dynamic formulation.



(a) (b)

Figure 2: (a) The spatial distribution of *pbc* and its components over the agent populations, and (b) the resulting statistical distribution. *pbc* is distinct from any of the three components, but the effects of each are visible. By allowing for more complexity than a simple measure, for example home value alone, *pbc* better captures the realities of solar economics, as related to affordability.

time, the payback is lower than the agent's *pbc*.[6] Therefore, if our *pbc* routine is robust, then at the time that the actual PV adopters acquired solar, their realized payback should have been lower than the *pbc* computed using Equation 3. As shown in Figure 3, the calculated *pbc* values resulted in 86.6% correct predictions for the $PP < pbc$ rule. This is an encouraging result, given that Equation 3 uses only basic publicly available information and there is no fitting associated with the choice of variables (I, s, W, and T) that go into computing *pbc* or with their functional forms. We believe that the remaining error in the model is due to the early stages of PV adoption in Austin (penetration level is less than 2%). Previous studies have found that at early stages of PV adoption there are indeed a segment of customers who adopt PV primarily for environmental reasons, even when the economics may be unattractive (for example, adopters with negative net present value) [82, 105]. As such, our model for computing *pbc* (Equation 3) is likely unable to account for these idiosyncratic adopters.

3.3.3 The Attitude Module: Combining Survey Data and Spatial Regression

Two dynamic, heterogeneous attributes – attitude (sia) and uncertainty about the attitude (U) – drive the attitudinal module. It is important to note that the value of these variables

[6] As discussed in Section 3.2, the economic criterion is a necessary but not a sufficient condition for adoption. Sufficiency requires both the social and the economic criteria to be met.

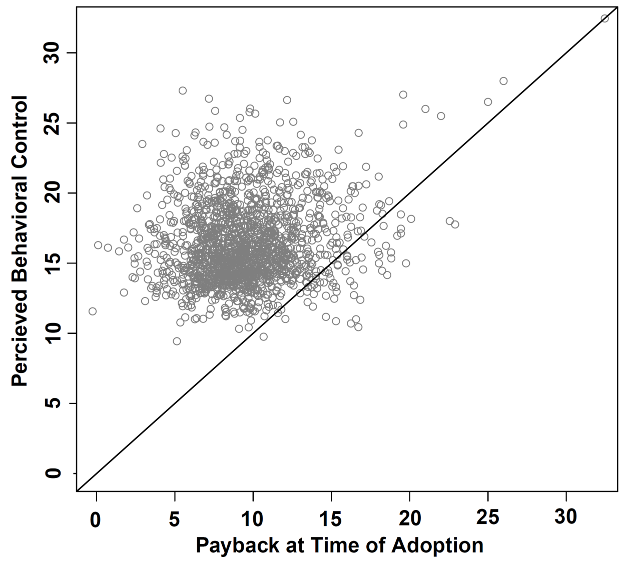


Figure 3: Estimates for each empirical adopter household's control measure *pbc* compared to their realized payback period PP in order to evaluate the initialization method in Section 3.3.1. The initialization method shown does much better than random, validating the decision rule for 86.6% of the empirical adopters.

for each agent varies over time, making the initialization process more difficult. Initial uncertainty U was distributed in proportion to the inverse absolute value of the initial sia for each agent. Our choice was based on previous behavioral research indicating that people are more likely to process relevant information and thus have higher uncertainty if they do not hold extreme attitudes [21, 67]. In the remainder of this section, we focus on the main attitudinal attribute itself, sia. For the model to be empirically grounded, the requirement of matching initial conditions must be addressed: at simulation initiation (time $t_0 = 0$, Q4 2007), sia needs to be initialized as close as possible to the *actual* sia at t_0 for all agents. In our model, this means every household in the study area. Measuring attitudinal variables necessitates self-reported questionnaires. Thus, obtaining the empirical *population-wide* distributions for heterogeneous agent attitudes using survey data would require surveying all households in the study area either longitudinally or on past beliefs about a technology (for example, perceptions of the profitability of solar). Longitudinal survey data collected since the early years of the adoption process for the technology of interest would be the ideal solution. Rarely is that data available; we believe that, although potentially the most accurate, this approach is not practical for most applications because of high costs and long lead times involved. The second approach would be a survey conducted in the present – potentially years after the beginning of the adoption process – that would ask agents (adopters and non-adopters) about their attitudes in the past (at t_0). While the adopters may still have better and more accurate recall of their attitude and other decision-variables at t_0, this approach is especially problematic for non-adopters, who, in general, are unlikely to have closely tracked their evaluation of a novel technology over time. As such, this approach is prone to significant measurement error. In general, then, both these approaches would result in high cost, unacceptable measurement errors, or both.

Instead, here we use a statistical modeling approach to derive population-wide estimates for agent attitudes at initialization, using attitude measurements on past beliefs for a sample of the adopters. The basic approach is to use kriging spatial autocorrelation model to take advantage of any spatial patterns in the data. Accordingly, sia is modeled and interpolated to the entire population in t_0 (Q4 2007), according to a three-step process:

Step 1: TPB posits that an individual's attitude towards a behavior arises from behavioral beliefs – beliefs of the individual about the likely outcomes of the behavior and her evaluation of those outcomes [2]. Following this reasoning, we create an index of survey items using data from 2004-2007 (Only respondents that were able to be geocoded were used: N = 108, i.e., 20% of the solar adopter population through 2007) using ten questions on the financial, environmental, and social aspects of their beliefs. Relevant survey questions related to these attributes are provided in *SI* (items 1-8). Critically, in this step, sia estimates are generated only for PV-adopter survey respondents (in Step three the relationships between these values and publicly available population data are used to generate estimates for the entire population). We further recognize that there is potential heterogeneity among agents in how much importance they place on each of the components of sia. For example, this allows for the situation where an agent may believe that solar is unprofitable, but profitability may not have been an important factor in that agent's decision to install solar. To account for this we use additional data from the PV-adopter survey to calculate a weighted average, where the weights (w_i) are the revealed importance of each component (relevant survey questions related to the weights are provided in *SI* (item 9). As noted above, this weighting is done only for the PV adopters for whom we have matched survey data (N = 108).

Thus, as shown in Equation 4, rather than being one simple measure of opinion regarding solar, sia for a given solar adopter i is an index of three components composed of financial (F_i), environmental (E_i), and social (S_i) belief indices for that adopter. The financial index F_i is the sum of survey respondent i's estimation of the payback period $\bar{PP}$, characterization of the profitability of the system Pr_i, and net monthly electricity bill savings Ms_i. The environmental index E_i is the sum of the level of overall environmental concern EC_i, the amount the individual is willing to pay to protect the environment $PayE_i$, and the level of concern for environmental issues in the individual's neighborhood $NeiE_i$.

$$sia_i = \frac{1}{3}(w_{1i}F_i + w_{2i}E_i + S_i) \ , \tag{4a}$$

$$F_i = \frac{1}{3}(\bar{PP}_i + Pr_i + Ms_i) \ , \tag{4b}$$

$$E_i = \frac{1}{3}(EC_i + PayE_i + NeiE_i) \ . \tag{4c}$$

We quantified social influence from neighbors as well as other acquaintances to account for the multiple interaction channels through which a potential adopter's attitude about solar are influenced. Accordingly, we took the social component S_i to be the average of two sub-indices:

$$S_i = \frac{1}{2}\left(w_{3i}\frac{\sum_{p=1}^{3} Nei_{pi}}{3} + w_{4i}A_{qi}\right), \text{where} \tag{5a}$$

$$log(N_i^{nei}+1), Mo_i, Cnf_i \in Nei_{pi} \text{ , and} \tag{5b}$$

$$A_{qi} = log(Ac_i + 1) \ . \tag{5c}$$

N_i^{nei} is the number of reported systems in the neighborhood, Mo_i is the degree of motivation obtained from neighborhood systems, and Cnf_i is the degree of confidence obtained from neighborhood systems. Ac_i is the number of contacts with PV owners outside the neighborhood. Note that N_i^{nei} and Ac_i are taken on the log scale to mitigate the influence of high counts in the index. Thus, S_i represents the contribution of social influence, including the influence of both neighbors and other acquaintances, in shaping an individual's attitude.

Step 2: As noted above, collecting past micro-data from the entire population of N households is infeasible and would likely involve a high degree of measurement error. Instead, to infer these values for each household we model the sia index created in Step one as follows:

$$sia_i = f(s_i, T_i/s_i, W_i/s_i) + \epsilon_i \ , \tag{6}$$

wherein sia is modeled as a function of home parcel size s, ratio of tree-cover to size (T/s), and home value per unit size (W/s), and ϵ is the error term. Note that the idea here is to express sia in terms of publicly known variables for every household in the study area. Due to the large number of potential covariates (up to fifth order polynomials were considered), model selection was performed via a stepwise procedure using the Akaike information criterion (AIC).

Step 3: While the model presented in Equation 6 provides a reasonable estimate of sia values, it assumes that there is no geographic relationship in agent attitudes beyond what may be captured by just s, T, and W. In reality, there may be unobserved attitudinal heterogeneity associated with geographical location for two reasons: first, geography may serve as a good proxy for additional socio-economic demographic variables not captured directly in our model, but which do impact attitudes (for example, education); and second, because attitudes may be expected to converge locally due to targeted marketing or neighborhood information exchange, for example through neighborhood associations or community organizations [77]. To account for this, we modify the model in Equation 6 using a spatial autocorrelation model (kriging with trend), giving:

$$sia_i = f(s_i, T_i/W_i, W_i/s_i) + m_i(x,y) + \epsilon_i^*(x,y) + \delta_i \ , \tag{7}$$

where ϵ_i^* is the kriging adjustment, m_i is the spatial trend adjustment, and δ_i is the residual error term accounting for variation in sia not explained by the publicly available data or geography. The function f describes the relationship obtained in the estimation of Equation 6. The kriging model acts as a spatial interpolator [42], using the sum of neighboring de-trended values, weighted by distance to location u $(= (x,y))$ and distance to nearby values, to estimate the spatial contribution to sia at locations where it is not directly observed.

3.3.4 Agent Attitude Initial Distribution

Based on AIC, the actual linear regression model that we estimate for Equation 6 using the survey data is as follows:[7]

$$\begin{aligned} sia_i &= \beta_0 + \beta_1 log(s_i) + \beta_2 log(s_i)^2 + \beta_3 log(s_i)^3 \\ &+ \beta_4(\frac{T_i}{s_i}) + \beta_5(\frac{T_i}{s_i})^2 + \beta_6(\frac{T_i}{s_i})^3 + \beta_7(\frac{W_i}{s_i})^2 + \beta_8(\frac{W_i}{s_i})^3 + \epsilon_i \ . \end{aligned} \tag{8}$$

Recall that in Step 1 *sia* was calculated for PV adopters between 2004-2007 with matched survey and appraisal district data. So, the coefficient values in Equation 8 are estimated using these households. The estimated model had an $AdjR^2$ of 0.21. Details of the kriging procedure and diagnostic plots for this model can be found in the *SI*.

As discussed in Step 3, before generating population estimates, we first adjust for spatial relationships to improve the *sia* model according to Equation 7. Figure 4a displays the total kriging adjustment values ($m_i + \epsilon_i^*$ in Equation 7) obtained by the kriging with trend model. The geographic distribution of standard errors around the kriging estimates are shown in Figure 4b. The green points on the map show the locations of pre-2008 adopters – these points were the precise locations to which the kriging model for spatial autocorrelation was fitted. As can be expected, standard error around ϵ^* increases around the edges of the study area and where survey samples were less abundant. Using the same set of adopters as used to estimate Equation 8, the kriging adjustment process increased the model $AdjR^2$ by an additional 0.15. Thus, the adjusted R^2 for the updated model given by Equation 7 was 0.36.[8] Finally, the *sia* estimates across the study area are obtained by using the estimated coefficients in Equation 7.

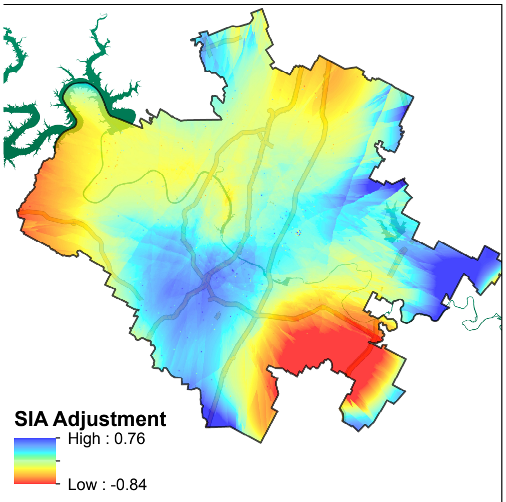
In sum, this technique shows how we have used multiple data-streams to address the initialization requirement, while avoiding the use of *ad hoc* random distributions. Future work could explore the application of the above approach for initializing a range of agent attributes in addition to *sia* as was done here.

3.4 Evolution of Agent Attitude

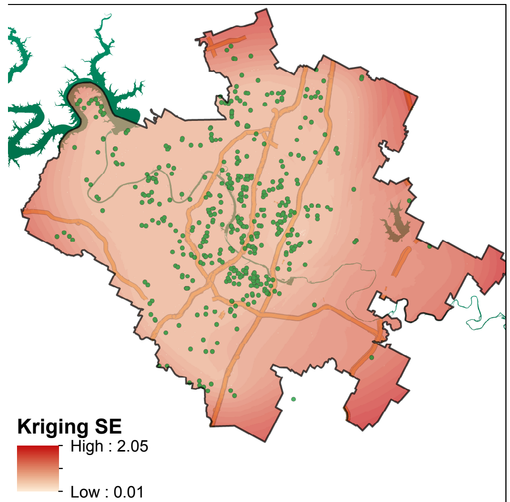
While the importance of modeling the *evolution* of agent attitudes is gaining recognition, many models still oversimplify or ignore this aspect [100]. Within ABM, the most common

[7] We used a backward selection process starting with 5th order polynomials and full interaction terms. It was in this process that the first order ratio term (W/s) was removed. Remaining variables shown are all significant.

[8] Overall, there are nine parameters in the *sia* model (Eq. 8) and six parameters in the kriging adjustment model (Eq. 8 in *SI*). Each of these two models is independently fitted using 108 data points, which is sufficiently large (observations-to-parameter ratio > 10) to mitigate any serious overfitting concerns; model diagnostics support this as well. As discussed in the *Validation* and *Results* sections, the integrated ABM model itself is independently validated using multiple criteria, including a "test" set of predictions through Q4 2014. To the extent overfitting in the *sia* model may be an issue, it would show up as poor validation metrics for the integrated model.



(a) Kriging adjustment ϵ^*



(b) Kriging standard error

Figure 4: (a) The spatial distribution of the full kriging adjustments (ϵ^*) used in Equation 7, and (b) the standard error around the kriging estimates as well at the locations of the pre-2008 solar adopter sample from the survey data (shown as green points).

models of attitude modification between agents are probabilistic [10], number-of-neighbors [26], percentage-of-neighbors [12], or simple averaging [1]. While easy to implement, these approaches oversimplify the reality of opinion dynamics as a complex, multi-dimensional process [19, 56, 63].

To better account for opinion dynamics in our behavioral model, at each time-step agents' attitudes about the technology (sia) and the uncertainties around those attitudes (U) are modified through interactions with other agents. Thus, as social norms are represented by the distribution of attitudes among other agents, they also are reflected dynamically in agent attitudes via agent-agent interactions. Interaction is modeled according to the Relative Agreement (RA) algorithm [23, 25, 48, 73]. In the RA algorithm, as the model moves forward though the simulated time-series, pairs of agents i and j interact, where i influences j. The extent to which such interactions alter agent j's attitude depends upon the overlap (the relative agreement) between agent i's and agent j's attitudes. In general, with the RA model agents are only influenced by other agents with relatively similar attitudes. A detailed formulation of the RA algorithm is presented in *SI*. At each time step, each agent interacts with ϕ other random agents from her social network. The choice of which agents interact is determined by the social network model: households are placed in small-world networks (SWN) where the majority of their connections are geographic and economic neighbors (live nearby and have similar wealth characteristics). Next we provide the details of how we use spatial and socio-economic demographic factors to construct the small-world network used in the solar ABM.

3.4.1 Agent Social Networks

Individual consumer attitudes are modified over time through social influence and interactions [114]. In the solar ABM we model this attitude evolution process to be the result of agent-agent interactions. Agent interactions with their connections (for example, friends and neighbors) depend on the social network structure. We use the small-world network model as the structure of the underlying social networks of agents in the solar ABM.[9] In the small-world network model, the definition of "local" needs to be resolved. For solar, contact with neighbors has been shown to drive down information costs, much more so than contact with non-neighbors [83]. This finding motivated a distance-based definition of local connections in the solar ABM. Accordingly, social network of agents in the model were largely made of households proximate in space to agent location.

The construction of agent networks was fully spatially resolved: actual agent locations and distances from other agents were used in generating the networks. Neighborhoods can be defined in a simple yet flexible manner by setting a radius r around each household. However, the choice of what r to use is not intuitive. In our model, r was determined by calculating a relevant distance from the empirical data: we looked at multiple distance bands, and calculated the spatial autocorrelation between adopter locations for each. In accordance with the known strong peer-effects in residential solar adoption, r was chosen as the distance at which PV adopter clusters are the strongest. Because these clusters arise through interaction within the neighborhood, the use of empirical clustering to determine the best distance by which to define the locals set arises from an observed pattern and prior empirical work. Importantly, this is an automated method that can be applied to any study area and for any technology influenced by social interaction effects.

Equation 9 shows how the level of correlation L for the study area A (in our case the city of Austin, Texas) is calculated as a function of distance d using a common transformation of Ripley's K function [28]. k is a weight assigned to a given pair of actual solar adopters i and j in Austin. For each value of d tried, k takes on the value of one if the distance between the points i and j in the GIS is less than or equal to d and zero otherwise. i_{rand} and j_{rand}

[9] A large body of research on the composition of social networks has shown that links between people can be categorized as mostly local (for example, geographically proximate) connections, with a minority of non-local connections [74, 92, 109]. These small-world networks have been used in ABMs to simulate the diffusion of technology through populations [24]. Empirical specification of social networks presents several challenges. The ideal approach for small, closed groups is individual network solicitation. For larger groups, solicitation based on random sampling is possible [6, 65]. However, for very large networks where explicit connection data is limited or costly, social network inference is possible using attributes such as proximity, wealth, and gender [76].

represent random points, used to generate the expected value $L(d)_{rand}$.

$$L(d) = \sqrt{\frac{A\sum_{i=1}^{n}\sum_{j=1}^{n} k_{i,j}}{\pi n(n-1)}} \tag{9a}$$

$$L(d)_{rand} = \sqrt{\frac{A\sum_{irand=1}^{n_{rand}}\sum_{jrand=1}^{n_{rand}} k_{irand,jrand}}{\pi n_{rand}(n_{rand}-1)}} \tag{9b}$$

$$r = max(L(d) - L(d)_{rand}) \tag{9c}$$

3.4.2 Agent Networks

We generated L scores for d values of 30.5 m (100ft) intervals up to 3050 m (10,000ft), and compared them to the expected (random) score L_{rand}. Effectively, this allowed us to test a multitude of distances to determine which was most relevant empirically, in the sense of strongest locational clustering of solar adopters. This process yielded a value of $r = 610$ m (2000ft). The static geographical network model with $r = 610$ m yielded an approximately normal degree distribution truncated at 0 with mean 498, and standard deviation 226.7. It is very unlikely that the average person will interact with 498 neighbors. To scale down the degree distribution to a more realistic range, we applied the additional constraint of wealth similarity. Homophily is a common attribute of social networks and has been found to play an important role in community structure [40, 49, 72]. In applying the homophily constraint we maintained proportionality, while reducing the locals-set by calculating the squared difference in wealth between the target node and its geographical neighbors: $(W_i - W_j)^2$; then, the 5% of agents with the smallest squared difference were connected as neighbors of i. Thus, geographic neighbors that are the most similar in wealth (as proxied by home value) were connected as locals. Finally, random connections were substituted for local connections with a 10% re-wiring probability to create a small-world structure; i.e., 10% of the local connections were replaced with random non-local connections with nodes anywhere in the population (the percentage of random connections is a parameter in the dynamic part of the model, sensitivities are shown in the *SI*). This yields a directional network. For example, it is possible that i could seek information from j, but j would not seek information from i. This property is common in information search and referral networks [16, 51, 62]. In the *SI* we show the resulting degree distributions and how the network created for the solar ABM compares to an equivalent Erdős-Rényi random graph.

3.5 Verification

After the ABM was constructed, the model operation was verified extensively through two parallel approaches: (i) testing of all sub-components using parameter sensitivity (sign and magnitude) as well as using simple test cases with known outcomes, and (ii) testing of the integrated model in a limited study area (one zip-code, results published in Robinson *et al.* [89]). Parameter sensitivity is reported in *SI*, available online. Reduction of the study area to one zip code allowed the model dynamics to be tracked through simple alterations in the agent decision rules by means of visualization of agent-states in real-time. For further verification, the model was slowed down by controlling time steps manually. At this spatial-temporal scale, the model was able to be run on desktop computing resources, and visualization was accomplished through the Agent Analyst Extension [54] in ESRI ArcGIS software.

3.6 Fitting

The model was fit to a real-world time-series of solar installation data from the solar program in Austin (see Section 3.1 for more details on data). Data from Q1 2004 - Q4 2007 were used for initialization, while data from Q1 2008 - Q2 2013 were used for fitting, yielding the "base case" integrated model (details described in Section 4.1). While there were several model outcomes (response variables) against which the model could be fit, *only the cumulative number of installations over time was used to fit the model parameters*. The deviation, or root mean squared error ($RMSE$) of the model was used as the objective function to minimize, and was calculated as follows:

$$RMSE = \sqrt{\sum_{q=1}^{n}(\frac{(\hat{a}_q - a_q)^2}{n})}\ , \tag{10}$$

where q is a given quarter, $\hat{a}$ is the cumulative number of predicted adopters by the model, and a is the number of cumulative adopters in the empirical data. The choice of cumulative installations as the fitting criterion serves several purposes. First, the total number of households that have installed solar at a given point in time is intuitive, and a measure tracked closely by policy-makers and program managers. Second, it is a highly aggregated measure, reducing the risk of over-fitting. Because the ABM operates at the individual level, fitting at the aggregate level allows the model to retain degrees of freedom because (over 2700 cumulative) adopters can be any households, anywhere in the study area, but still meet the fitting criterion. This type of error will only show up in validation, not fitting. This allows for validation along multiple *unfitted* outcomes, thereby providing a more powerful test of the predictive capability of the model. Finally, fitting to the cumulative adoption level means that our validation criteria will have very comparable sample sizes, increasing the robustness of the validation metrics, which are discussed below in Section 3.7.

Six structural parameters were used to specify and control the social networks, opinion convergence, and the distribution of the *control* variable (pbc). ϕ controls the number of interactions per agent per quarter. μ is the coefficient of convergence in the RA algorithm (see Appendix). λ^r controls the percentage of random connections in each agent's social network. α_0 and α_1 define the intercept and slope used in the linear scaling procedure for calculating pbc. sia^{thresh} is the global attitude threshold value necessary for a household to become an adopter. A global threshold allows for the exchange of attitudes to occur at the same scale, and it also places the model in the broader context of threshold models [104].[10] Starting with reasonable guesses for the values of the parameters to be fit, the fitting of the

[10] There are three primary reasons for using a global threshold for sia. 1) Practically, because attitude

model parameters was done iteratively. Before the fitting process began, an initial reasonable guess for each parameter was generated using prior work [82–84, 89] and/or pre-fitting test runs to observe the behavior of the model in response to different sets of parameter values. Then, each model parameter was varied systematically along a set range while the others were held constant. The parameter value that generated the minimum RMSE (cumulative) was selected and held, while the next parameter was varied. After a full parameter sweep, the process was repeated two more times, at which point the marginal decrease in RMSE by further adjustment was less than 2%. Sensitivity testing on these parameters is reported in the *SI*.

Because there was some randomness in the simulation (for example in the order in which the agents act, and which subset of connections are chosen randomly in the RA algorithm), batches of 100 different runs were used in generating the statistics of the variables of interest that were used in fitting and validation. For the RMSE calculation (Eq. 10), residuals for the cumulative number of adopters were calculated by subtracting each model output for a given quarter from the empirical (actual) outcome in the same quarter.

3.7 Validation

We validate the model across four model outcomes: predictive accuracy; RMSE of marginal adoptions temporally; spatial accuracy; and demographic accuracy. Besides providing evidence of the model's adequacy in representing the target system (solar technology adoption), the emergent properties associated with each validation metric are linked to important policy or infrastructure planning questions and are therefore interesting in their own right.

3.7.1 Validation of Predictive Accuracy

A predictive forecast was run using the base-case model parameters on a test set of data held back during the fitting process. Recall that the model was fit using data from twenty-two quarters (Q1 2008-Q2 2013). The predictive model uses the parameter values fitted to this data to forecast out six quarters, Q3 2013 to Q4 2014. Predictions were compared to empirical adoption levels over this period, which was not used in any fitting process.

3.7.2 Temporal Validation

The instantaneous rate at which adoption is occurring at a given point in time is the slope of the cumulative adoption curve, and at a marco-level it shows whether the technology

is "traded" between agents in the Relative Agreement algorithm, it is important that these trades occur on an equal units basis. Individual threshold would greatly impact the value of a given exchange, and thus the attitude scale. 2) Theoretically, a global *sia* threshold fits with the original formation of opinion dynamics models, including Relative Agreement. Furthermore, this mimics threshold models, which are fairly common in the literature and well understood. 3) Finally, individual thresholds would require additional fitting mechanisms to assign those thresholds to individuals to begin with, thereby making the model more complex, while significantly reducing the degrees of freedom.

is diffusing more slowly or more rapidly. We validate the rate of adoption (the number of *new* adopters in quarter q) in our models using the RMSE (Equation 10) for the *marginal* number of adoptions each quarter. While the marginal adoption rate is related to the fitting metric (the cumulative adoption over time), due to variations in prices and rebates in the study area it is not entirely dependent on the cumulative number of installations. Thus, the marginal RMSE is an admissible independent metric for external validation.

3.7.3 Spatial Validation

We validate the geographic distributions of solar adopters across the study area according to three different statistics based on the density of systems per square mile. *It is important to note that these statistics were not used in model fitting* (Section 3.6). The methodology for calculating error over space has received quite a bit of attention in the geography and remote sensing literature [80, 107]. While most statistical comparison of maps relies on arithmetical cell-by-cell evaluation [110], this method can be flawed for many applications because it ignores the spatial structure of the errors. We report the simple arithmetic error ($empirical_i - predicted_i$, where i is a cell in a spatial grid or raster) as well as two additional measures to evaluate the spatial prediction errors: fuzzy numerical similarity (κ^*) and wavelet verification (r^w). The first step was to create an adoption probability for each agent in the model by averaging over the simulation outcomes across all the 100 runs in a given batch.[11] Next, a Gaussian kernel density function[12] with the predicted adoption probabilities as inputs was used to calculate the number of systems per square mile over the entire study area at 100ft raster resolution. This yielded the *simulated raster*, used for computing the fuzzy numerical and the wavelet verification metrics. The same function was used over the empirical data as the simulated data, allowing the two maps to be quantitatively compared and spatial error to be calculated.

One benefit of κ^* is that it rewards local similarity. Further details of the calculation of κ^* are provided in *SI*. While the fuzzy numerical method is useful for assessing spatial similarity, the parameterization of the smoothing function can impact the obtained metrics. In order to further check for robustness, we calculated correlation coefficients using wavelet verification. Wavelet verification has gained popularity in the meteorological literature due to the need to compare forecasts against observed weather patterns at different resolutions [15, 18]. 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^w) can be generated. The discrete wavelets aggregate the rasters to coarser resolutions [15]. In this study Harr wavelets and 8^{th} level aggregation (8x8) were used.

[11] A 'batch' refers to all the 100 independent runs of a specific model with the same set of parameter values.

[12] This was accomplished using the KernelDensity() function in ESRI's Spatial Analyst ArcPy library.

3.7.4 Demographic Validation

In our study area, as of Q2 2013, the average non-adopter home value was $267,965.80, compared to the average adopter home value of $475,326.40. However, as the costs of solar decline over time, one would expect to see the technology being adopted by less wealthy households. This downward trend was reflected in the empirical data used in this study, and thus should be reproduced as an emergent property by successful models. RMSE was calculated (Equation 10) for comparison of predicted and empirical outcomes, where $\hat{a}$ was the median adopter home value for a given quarter q in the model and a was the empirically observed median adopter home value for that quarter.

As a further step, we validated adopter home values over space as well. This served three purposes: first, it is plausible that the model could meet the aggregate demographic validation criterion presented above but still be inaccurate at more resolved spatial scales (for example, by predicting too-wealthy households in some neighborhoods, but performing well on the city-level median). Second, it aligns the spatial validation criterion with the explanatory scale (neighborhoods, roughly 610 m in radius) of the model. Third, by clearly pointing to areas of relatively higher demographical error in the model, this validation could help identify potentially influential variables for further improving the model. We used a similar methodology as described in Section 3.7.3 to calculate the simple error, fuzzy numerical similarity, and wavelet correlation coefficients. It is worth noting that results from the two metrics (spatial adoption density versus home value of adopters) are highly differentiated: for example, shifting adopters just one house to the north would have almost no visible effect on the density of systems per square mile, but could alter the adopter home values considerably. This point is further explained in *SI*.

4 Results

4.1 The Integrated Model

In this section we present the results of fitting and validation of the integrated model (also, the "base case"). The integrated model is the full model that uses all model components as described in Section 3.2 (Figure 1). The integrated model also uses empirical distributions to initialize all agent states (Section 3.3). As such, it is fully empirically grounded in terms of the description of the system, in the sense that all model components are initiated and populated with the most granular data available in the study. Integrated Model fit (RMSE in cumulative adoptions, Figure 5) was found to be low, at 117.81. As RMSE units are always the same as the quantity being estimated, this shows that on average the Integrated Model was off by 117 households. For comparison, this is just 0.06% of the 173,466 potential adopters in the study area, or 4.3% of the cumulative number of adopters in Q2 2013 ($N = 2,738$). This is not surprising given the strong empirical foundation of the model. Recall that model fitting is the basis for parameter value selection, and is done only on the RMSE in the cumulative installation levels over time (Section 3.6). The obtained parameter values are listed in Table 2.

Table 2: Parameters active in the Integrated Model and the values obtained during the fitting process described in Section 3.6.

Model Parameters in the Integrated Model					
ϕ	μ	χ	α_0	α_1	sia^{thresh}
4.0	0.38	0.1	-60.61	2.46	0.6

4.1.1 Predictive Accuracy

In order to test the model's predictive validity, the model was run forward through Q4 2014, as described in Section 3.7. These predictions were compared to historical data over the same period (Figure 5). We emphasize that no additional fitting was performed. The unfitted test set from Q3 2013 to Q4 2014 made up 21.4% of the data (by number of quarters). The RMSE in cumulative installations for the unfitted adoption forecast was 192.3 (households). This is 5.5% of the empirical cumulative adoption at the end of Q4 2014 ($N = 3488$). Importantly, as can be seen in Figure 5 the forecast is able to replicate the flattening of the empirical curve that starts around Q4 2013 and the subsequent upswing in adoption. This kind of behavior is quite difficult to replicate using parametric forecasts (Basic Structural Models, ARIMA models) without clear seasonal patterns in the historical data. These results provide strong support for the predictive capability of our model.

4.1.2 Temporal, Spatial, and Demographic Validation

The three additional validation metrics discussed in Section 3.7 for the integrated model are reported in Table 3. Overall, validation of the integrated model was highly supportive of the model structure, components, and parameter values as being a strong description of the target system and showed little evidence of over-fitting. Temporal validation of the integrated model using marginal adoption RMSE showed only one period of large error, in Q4 2011.[13] As shown in Table 3, the spatial distribution of adopters predicted by the integrated model was very similar to that seen in the empirical data. Among a number of other model variations that were tested[14], the integrated model showed the lowest average spatial arithmetic error (0.46), highest fuzzy numerical κ^* (0.43), and strongest wavelet correlation r^w (0.86). While there were still some areas of under- and over-prediction (blue and red in Figure 6, lower panel), the structure and location of most PV dense areas (shown in darker brown in Figure 6) matched well. The model also performed well demographically,

[13] In our dataset in Q4 2011 the dollar per Watt installed costs of solar decreased substantially. The price decline results in a surge in installations in the model greater than that observed in the empirical data. With this quarter removed, the Integrated Model RMSE is much lower, at 45.79.

[14] These model variations were created by systematically reducing the empirical basis of the model and/or simplifying agent decision rules. The full findings of the systematic model comparisons are in preparation as a separate manuscript. Results are available upon request.

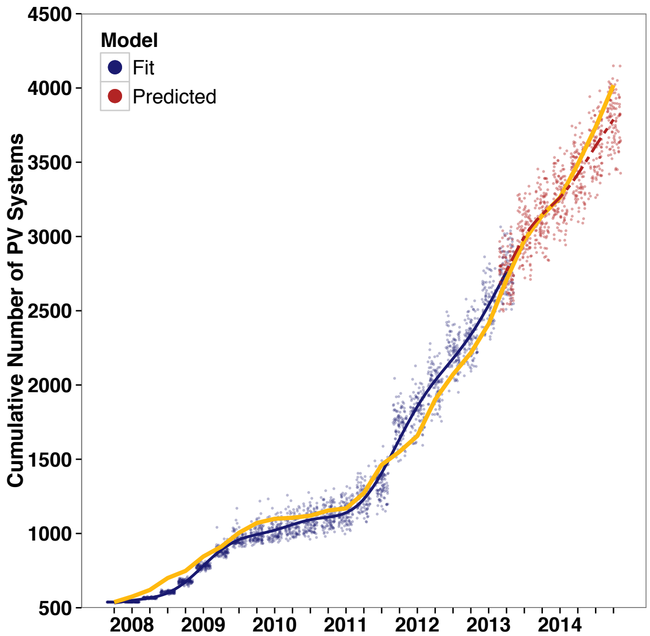


Figure 5: Fit and predicted model outcomes compared to empirical data. Observed adoption levels are shown by the gold line. The purple line shows the Integrated Model, described in Section 4.1. The average of the six quarter forecast (Q3 2013 - Q4 2014) is shown as the red dashed line (Section 4.1.1). The points represent individual model runs. No parameters were altered or fitted to data after Q2 2013.

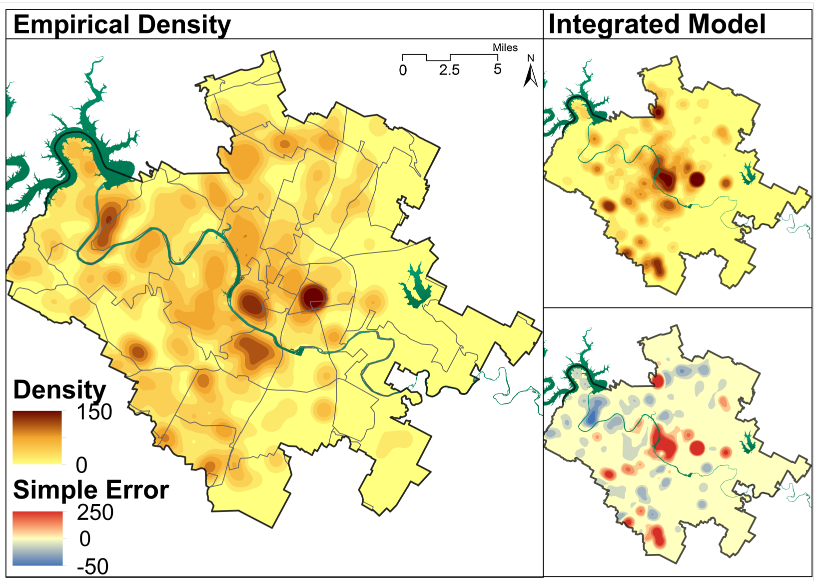


Figure 6: Spatial system-density (systems per square mile or 1.61 square kilometer) validation for the integrated model. The empirical system density is shown in the far left panel. The top (yellow-brown) panels show simulated system density, while the lower (divergent blue-red) panels show simple error (Empirical - Simulated) in system density. Raster resolution is 30.5 m. Additional spatial validation metrics, fuzzy numerical κ, and r^w are reported in Table 3.

and of the various models tried it had the lowest error with regard to predicted adopter home values (Table 3 and Figure 7). Further, as reflected in the low demographic RMSE, the integrated model matched the overall slow downward trend in adopter wealth over time (not shown). Although there are still spaces where this model shows non-negligible spatial demographic error – particularly two areas in North and West Austin where the model is predicting home values of adopters to be lower than observed – overall the integrated model performs quite well in spatial demographic validation (Figure 7). This suggests that not only does the model represent average adopter home values well, but that it can account for local neighborhood variations as well.

Table 3: Summary of validation results for the four models. Descriptions of each of the metrics can be found in Section 3.7. Temporal, spatial, and demographic validation metrics are independent of the fitting criteria.

Validation Metrics for the Integrated Model							
RMSE marginal	**Sp: Simple**	**Sp: Fuzzy κ**	**Sp: r^w**	**Dem: RMSE**	**Dem: Simple**	**Dem: Fuzzy κ**	**Dem: r^w**
76.93	0.46	0.43	0.86	110,580.2	-37,967.7	0.81	0.81

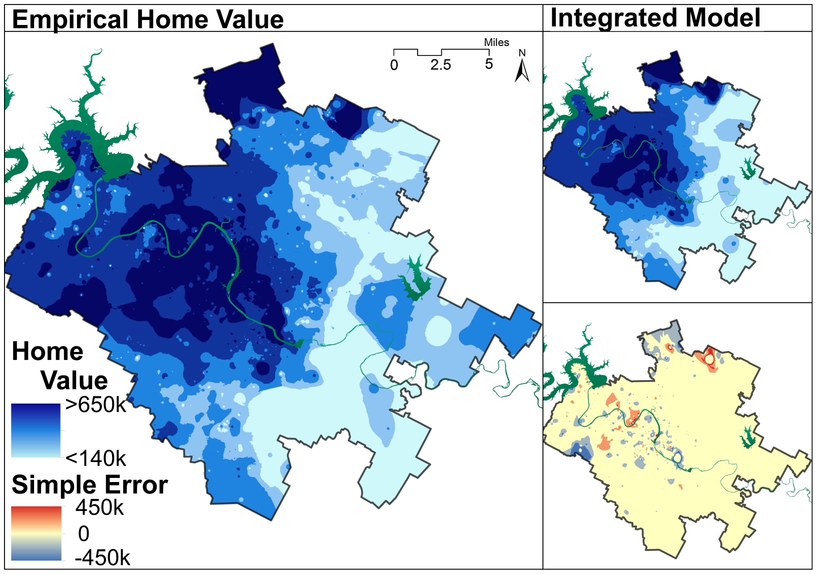


Figure 7: Spatial demographic validation for the integrated model. The large panel on the left shows the empirical adopter home-value distribution over space. The top (blue-scale) panel show simulated adopter home values. Adopter home values were interpolated using the Inverse Distance Weighted technique (see *Supplementary Information*). The lower (divergent blue-red) panel show simple error (Empirical- Simulated). Raster resolution is 30.5 m. Additional spatial validation metrics, RMSE, fuzzy numerical κ, and r^w are reported in Table 3.

4.2 Applications

The development of a validated integrated model allows for virtual policy experiments to be performed. The flexibility around these application scenarios is a major strength of ABM. Below we present two illustrative applications of the ABM model developed here (see Section 1.2 for the motivation behind these applications.).

4.2.1 Addressing Equity Concerns: Additional Rebates for Low-Income Households

Additional rebate offerings for low-income households were simulated in two scenarios using the the Integrated Model (Section 4.1). Rebate levels were increased by \$0.2 per Watt in the first scenario and \$0.4 per Watt in the second scenario. In this simulation experiment, these *additional* rebates were offered to only those households in the bottom quartile of wealth (proxied by home value). The additional rebates were offered to low-income households for the entire simulation period. To assess the impact of additional rebates, the change in adoption among low-income households was measured in the simulation experiment relative to the unaltered Integrated Model results presented in Section 4.1. Results are shown in Table 4. In the Integrated Model 2.5% of adopter households fall in the low-income category. In the \$0.2 per Watt rebate increase scenario, the proportion of low-income adopters increases to 3.3%. In the \$0.4 per Watt rebate increase scenario, the proportion of low-income adopters increases to 4.4%. While these results are not dramatic in absolute terms, the relative increase over baseline levels is substantial (33% and 81.5%, respectively). The relatively low impact (in absolute terms) of the simulated rebate increases suggests that large rebate increases may be needed to drive low-income solar adoption.

Table 4: Results of low-income solar rebate scenarios.

Low-Income Solar Rebates			
Model	Average Q2 2013 Adopters	Median Adopter Home Value ($)	Low-Income Adopters
Integrated Model	2,760.6	374,402	2.5%
+0.2$/Watt	2,784.5	373,146	3.3%
+0.4$/Watt	2,798.4	372,085	4.4%

4.2.2 Impact of Rebate Level Changes on Adoption

A question that solar program designers often face is how much would adoption change if rebate levels were changed. To examine this question we performed a simulation experiment wherein the price of solar to the customer ($/Watt, net of any rebates and incentives) was systematically varied compared to historical observed prices, which were used in the Integrated Model (Section 4.1). To mimic increased rebate levels, we created 10 scenarios. In each scenario, prices (faced by the customer) were decreased in steps of 0.025 $/W up to a maximum of 0.25 $/W; so in the first scenario the price decrease (P) was 0.025 $/W and in the tenth scenario it was 0.25 $/W. The additional rebate in each scenario remained in effect throughout the simulation period, Q1 2008 through Q2 2013. 100 model runs were performed for each scenario. The impact on adoption was obtained by regressing Δ, the change in quarterly adoption relative to the base case on the exogenous change in price P and the lagged (i.e., the prior quarter) cumulative number of adopters Q_{it-1} in the scenario, where i denotes a specific price-change scenario and t is a quarter:

$$\Delta_{it} = \beta_0 + \beta_1 P_i + \beta_2 Q_{it-1} + \epsilon_{it}. \tag{11}$$

A model with an interaction between P and Q was also performed:

$$\Delta_{it} = \beta_0 + \beta_1 P_i + \beta_2 Q_{it-1} + \beta_3 P_i \times Q_{it-1} + \epsilon_{it}. \tag{12}$$

The pooled regression results combining all 10 scenarios together are shown in Table 5. The simple model (Eq. 11) shows that on average holding the number of adopters constant, a $0.1/W decrease in the price of solar PV for the customer increases adoption by about 10 *additional* (i.e., over the base case) new adopters each quarter in the study area.[15] That

[15]Note that the coefficient estimates correspond to a $1/W decrease in price. So to get the effect for a $0.1/W the estimates need to be divided by 10.

is equivalent to a 8% increase in adopters over the full simulation period compared to the base case. In the interaction model (Eq. 12), the estimate for price alone reduces to about 5 additional new adopters for a \$0.1/W price decrease. The interaction effect between price and the installed base is strong and significant ($\beta_3 = 0.028$). This suggests that rebate changes impact adoption levels more strongly further down in the solar program, when the installed base is more substantial. For example, near the beginning of the program, when installed capacity is quite low, a \$0.1/W price decrease increases quarterly adoptions by an additional 5 adopters. But the impact is magnified later in the program: when the installed base reaches 1000, a \$0.1/W price decrease leads to an additional 8 new adopters per quarter ($5 + 0.0028 \times 1000$) , and the same effect at an installed base of 3000 is about an additional 13 new adopters per quarter. This intuitively makes sense because in our model an agent needs to be both socially and economically activated to adopt solar. *Ceteris paribus*, since at a larger installed base non-adopter agents have more opportunities to interact with existing adopters and hence becoming socially activated, an increase in rebates will have more impact because the more attractive economics now can operate on a broader section of the non-adopters that is socially activated.

Table 5: Impact of rebate/price changes on the level of PV adoption.

	Simple Model		**Interaction Effects**	
Coefficient	Estimate	p-value	Estimate	p-value
Intercept: β_0	-5.53	<0.0001	-1.28	0.059
Price: β_1	97.44	<0.0001	50.58	<0.0001
Installed Base: β_2	0.003	<0.0001	0.001	0.011
Interaction: β_3			0.028	<0.0001
Adj R^2	0.064		0.068	

5 Conclusion

With the goal of developing models that are capable of robustly representing the bounded rationality of individual decision-makers, in this paper we presented the architecture of a theoretically-based and empirically-driven agent-based model of technology adoption, with an application to residential solar PV. We focus on the theoretical and empirical aspects of model design, setup, initialization, and validation. Our main emphasis was to attempt to address two major concerns with the use of ABM in human-technical systems: poor integration of data in model initialization and validation, and *ad hoc* definitions of agent behavioral rules. Toward that end, we overlaid multiple data streams covering 2004-2013 including survey, program, appraisal district, and LIDAR data to set up an empirical ABM bound closely to the target system – namely, the adoption of residential solar PV at the city scale. Driven by the Theory of Planned Behavior, agents' adoption decisions are jointly determined by both attitudinal and control beliefs. Both the attitudinal and control sub-models are dynamic, with the key metrics evolving depending on market conditions or via social interactions within agent networks.

For the control sub-model, we combined publicly available data on home value, size of the home, tree-cover, and irradiance (amount of sunlight) to develop the control variable representing the agent's perception of whether she could afford solar or not, when compared with a simple time-resolved payback calculation. Using this metric, the simple rule: "An agent becomes economically activated (able to install solar, given a favorable attitude toward the technology) when control is greater than payback" was found to be consistent with 86.6% of the real adopters in the empirical data. We believe this is an encouraging result since it uses only publicly available data and no fitting is involved in the selection of the variables that go into developing the control metric.

We also presented a technique to generate population-wide estimates for agent attitudes through regression of survey data on population variables, generated from publicly available datasets. We further improved these estimates using a kriging model, where we take advantage of geographically correlated, but often unobservable variables such as education, familial composition, retirement status, race, and political affiliation. Using only publicly available data, this technique explained 36% of the variance in weighted solar adopter attitudes regarding the financial, environmental, and social aspects of solar ownership (as measured in a survey). Since the approach of this technique is general, it can be used for other applications, and it significantly decreases the time and cost associated with population-scale empirical ABM by reducing the need for expensive and time-consuming survey-based data collection.

Further, in line with existing literature we find that small-world social networks where locals are based on geography alone, even when the relevant distance is derived from empirical patterns, generated very high degree distributions. To achieve more realistic degree distributions, we further refined these connections based on home-value similarity. This approach affords two benefits. First, it maintains proportionality: agents in the dense neighborhoods identified in the eigenvector centrality comparison will still have more connections than agents in sparse neighborhoods. Second, unlike decreasing the degree distribution by choosing a smaller radius, it allows for observed empirical patterns to be maintained. In the case of solar adoption in the study area, this pattern was the finding that the greatest correlation between solar owner locations occurs at about 610 m radial neighborhoods.

The resulting Integrated Model was fit to the cumulative number of quarterly adoptions using six parameters. The model fit was very good, especially considering the low proportion of solar adopters in the population. Importantly, the model replicated the major structural features in the empirical diffusion curve. However, alone this fit would not be sufficient for establishing the model's performance: indeed, a relevant criticism of empirical ABM is the potential for over-fitting. To overcome this, the Integrated Model was validated against three external criteria, using five different metrics. In other words, we cross-validated the model against five other outcome variables independent of the fitting criteria. Finally, we also compared predictions generated by the model to a portion of data not used in fitting (the test set, Q3 2013 - Q4 2014). The integrated model performed well in each of these validation cases, outperforming other fitted models across the board.[16]

[16]A systematic exploration of different models and quantification of the importance of each sub-model

Two policy applications tested in this paper show how the developed ABM framework can be used as a virtual laboratory. First, by simulating a low-income solar program within the context of the fitted model, we provide evidence that rebate levels must be quite high to have a large absolute impact on adoption of solar PV by low-income households. Second, by simulating changes in the rebate-level we find that the effect of a change in the rebate on PV adoption scales with the installed base: early in the program when there are few adopters, changes in PV rebate levels have modest impact, while the impact is much greater later on because of social effects.

Overall, in this paper we have presented a comprehensive approach for the integration of granular and overlapping data-streams to ground ABM of energy technology adoption empirically, while building upon theoretical underpinnings. Further, in order to address concerns regarding representativeness and over-fitting in ABM generally, we have developed and applied a multi-pronged external validation process to the integrated agent-based model. This is important because only by thoroughly grounding the model components in real-word data and through rigorous validation can ABM generate relevant policy insights, predictions, and emergent properties beyond the reach of conventional models.

Acknowledgements

We thank Benjamin Picone and Abhishek Tondon for research assistance and Jay Zarnikau and Eugenio Arima for very helpful suggestions on earlier versions of this work. We acknowledge funding support from the U.S. Department of Energy under its Solar Energy Evolution and Diffusion Studies (SEEDS) program within the SunShot Initiative (Award Number DE-EE0006129). Data assistance from Austin Energy and the City of Austin is gratefully acknowledged. Computing resources made available by the Texas Advanced Computing Center (TACC) were critical for this work and are gratefully acknowledged. Insightful and constructive feedback from four anonymous reviewers was very helpful in improving the manuscript. VR acknowledges support from the Elspeth Rostow Memorial Fellowship and from the Policy Research Institute (PRI) at the LBJ School of Public Affairs (UT Austin). All remaining errors are ours alone.

References

[1] D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi. Spread of (mis)information in social networks. *Games and Economic Behavior*, 70(2):194–227, 2010.

[2] I. Ajzen. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211, 1991.

in model fit and validation is in preparation as a separate manuscript, and the results are available upon request.

[3] I. 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.

[4] L. An. Modeling human decisions in coupled human and natural systems: Review of agent-based models. *Ecological Modelling*, 229(0):25–36, 2012. Modeling Human Decisions.

[5] C.J. Armitage and M. Conner. Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, 40(4):471–499, 2001.

[6] J. Axsen, C. Orlebar, and S. Skippon. Social influence and consumer preference formation for pro-environmental technology: The case of a U.K. workplace electric-vehicle study. *Ecological Economics*, 95(0):96–107, 2013.

[7] R.M. Baron and L.A. Boudreau. An ecological perspective on integrating personality and social psychology. *Journal of Personality and Social Psychology*, 53(6):1222, 1987.

[8] F.M. Bass, T.V. Krishnan, and D.C. Jain. Why the bass model fits without decision variables. *Marketing Science*, 13(3):203–223, 1994.

[9] S.T. Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262, 1994.

[10] S.C. Bhargava, A. Kumar, and A. Mukherjee. A stochastic cellular automata model of innovation diffusion. *Technological Forecasting and Social Change*, 44(1):87–97, 1993.

[11] G. Blackburn, C. Magee, and V. Rai. Solar valuation and the modern utility's expansion into distributed generation. *The Electricity Journal*, 27(1):18–32, 2014.

[12] 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.

[13] B. Bollinger and K. Gillingham. Peer effects in the diffusion of solar photovoltaic panels. *Marketing Science*, 31(6):800–812, 2012.

[14] M. Brede and B.J.M. De Vries. The energy transition in a climate-constrained world: Regional vs. global optimization. *Environmental Modelling & Software*, 44:44–61, 2013.

[15] W.M. Briggs and R.A. Levine. Wavelets and field forecast verification. *Monthly Weather Review*, 125(6):1329, 1997.

[16] J.J. Brown and P.H. Reingen. Social ties and word-of-mouth referral behavior. *Journal of Consumer Research*, 14(3):350–362, 1987.

[17] S. Cantono and G. 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. Evolutionary Methodologies for Analyzing Environmental Innovations and the Implications for Environmental Policy.

[18] 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.

[19] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591, 2009.

[20] T.L. Chartrand and J.A. Bargh. The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6):893, 1999.

[21] 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.

[22] J. Conlisk. Why bounded rationality? *Journal of Economic Literature*, 34(2):669–700, 1996.

[23] 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.

[24] 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.

[25] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(1):87–98, 2000.

[26] 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.

[27] T. Dietz, P.C. Stern, and E.U. Weber. Reducing carbon-based energy consumption through changes in household behavior. *Daedalus*, 142(1):78–89, 2013.

[28] P.M. Dixon. Ripley's K function. In *Encyclopedia of Environmetrics*. John Wiley & Sons, Ltd, 2006.

[29] C. Dong. *Technology diffusion policy design: cost-effectiveness and redistribution in California solar subsidy programs*. PhD thesis, The University of Texas at Austin, 2014.

[30] S.N. Durlauf. Complexity, economics, and public policy. *Politics, Philosophy & Economics*, 11(1):45–75, 2012.

[31] F. Eggers and F. Eggers. Where have all the flowers gone? Forecasting green trends in the automobile industry with a choice-based conjoint adoption model. *Technological Forecasting and Social Change*, 78(1):51–62, 2011.

[32] J. M. Epstein. Why model? *Journal of Artificial Societies and Social Simulation*, 11(4):12, 2008.

[33] 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.

[34] G. Fagiolo, A. Moneta, and P. Windrum. A critical guide to empirical validation of agent-based models in economics: Methodologies, procedures, and open problems. *Computational Economics*, 30(3):195–226, 2007.

[35] G. Feola and C.R. Binder. Towards an improved understanding of farmers' behaviour: The integrative agent-centred (IAC) framework. *Ecological Economics*, 69(12):2323 – 2333, 2010.

[36] 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.

[37] R. Gelman and M. Meshek. 2012 renewable energy data book. Technical report, U.S. Department of Energy, 2013.

[38] M.D. Gerst, P. Wang, A. Roventini, G. Fagiolo, G. Dosi, R.B. Howarth, and M.E. Borsuk. Agent-based modeling of climate policy: An introduction to the engage multi-level model framework. *Environmental Modelling & Software*, 44:62–75, 2013.

[39] G. Gigerenzer and R. Selten. *Bounded rationality: The adaptive toolbox*. MIT Press, 2002.

[40] M. Girvan and M.E.J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[41] 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.

[42] P. Goovaerts. *Geostatistics for Natural Resources Evaluation*. Oxford university press, 1997.

[43] M. Graziano and K. Gillingham. Spatial patterns of solar photovoltaic system adoption: the influence of neighbors and the built environment. *Journal of Economic Geography*, doi:10.1093/jeg/lbu036, 2014.

[44] P.E. Green and V. Srinivasan. Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, 5(2):103–123, 1978.

[45] V. Grimm, U. Berger, D.L. DeAngelis, J. G. Polhill, J. Giske, and S.F. Railsback. The ODD protocol: A review and first update. *Ecological Modelling*, 221(23):2760–2768, 2010.

[46] 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.

[47] M. Gunther, C. Stummer, L. Wakolbinger, and M. Wildpaner. An agent-based simu-

lation approach for the new product diffusion of a novel biomass fuel. *Journal of the Operations Research Society*, 62:12–20, 20011.

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

[49] A.D. Henry, P. Pralat, and C. Zhang. Emergence of segregation in evolving social networks. *Proceedings of the National Academy of Sciences*, 108(21):8605–8610, 2011.

[50] A. Heppenstall, A. Evans, and M. Birkin. Using hybrid agent-based systems to model spatially-influenced retail markets. *Journal of Artificial Societies and Social Simulation*, 9(3), 2006.

[51] O. Hinz and M. Spann. The impact of information diffusion on bidding behavior in secret reserve price auctions. *Information Systems Research*, 19(3):351–368, 2008.

[52] T. Islam. Household level innovation diffusion model of photo-voltaic (PV) solar cells from stated preference data. *Energy Policy*, 65(0):340–350, 2014.

[53] T. Islam and N. Meade. The impact of competition, and economic globalization on the multinational diffusion of 3G mobile phones. *Technological Forecasting and Social Change*, 79(5):843–850, 2012.

[54] K.M. Johnston, M.J. North, and D.G. Brown. *Agent Analyst: Agent Based Modeling in ArcGIS*. ESRI Press, Redlands, California, 1 edition, 2013 2013.

[55] P. Kaufmann, S. Stagl, and D.W. Franks. Simulating the diffusion of organic farming practices in two new EU member states. *Ecological Economics*, 68(10):2580–2593, 2009.

[56] H.C. Kelman. Processes of opinion change. *The Public Opinion Quarterly*, 25(1):57–78, 1961.

[57] R. Kemp and M. 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.

[58] A.P. Kirman and N.J. Vriend. Evolving market structure: An ACE model of price dispersion and loyalty. *Journal of Economic Dynamics and Control*, 25(34):459–502, 2001. Agent-based Computational Economics (ACE).

[59] D. Kniveton, C. Smith, and S. Wood. Agent-based model simulations of future changes in migration flows for Burkina Faso. *Global Environmental Change*, 21(S1):S34–S40, 2011.

[60] C.E. Laciana and S. 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, 2011.

[61] 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.

[62] J. Leskovec, A. Singh, and J. Kleinberg. Patterns of influence in a recommendation network. In *Advances in Knowledge Discovery and Data Mining*, pages 380–389. Springer, 2006.

[63] J. Lorenz. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12):1819–1838, 2007.

[64] 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.

[65] A. Maertens and C.B. Barrett. Measuring social networks' effects on agricultural technology adoption. *American Journal of Agricultural Economics*, 95(2):353–359, 2013.

[66] 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.

[67] 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.

[68] 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.

[69] R.B. Matthews, N.G. Gilbert, A. Roach, J.G. Polhill, and N.M. Gotts. Agent-based land-use models: a review of applications. *Landscape Ecology*, 22(10):1447–1459, 2007.

[70] E. Mazhari, J. Zhao, N. Celik, S. Lee, Y.J. Son, and L. Head. Hybrid simulation and optimization-based design and operation of integrated photovoltaic generation, storage units, and grid. *Simulation Modelling Practice and Theory*, 19(1):463–481, 2011.

[71] D. McFadden. Economic choices. *American Economic Review*, 91(3):351–378, 2001.

[72] M. McPherson, L. Smith-Lovin, and J.M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.

[73] M. Meadows and D. Cliff. Reexamining the relative agreement model of opinion dynamics. *Journal of Artificial Societies and Social Simulation*, 15(4), 2012.

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

[75] 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.

[76] K. Munshi. Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Journal of Development Economics*, 73(1):185–213, 2004.

[77] D. Noll, C. Dawes, and V. Rai. Solar community organizations and active peer effects in the adoption of residential PV. *Energy Policy*, 67:330–343, 2014.

[78] M.J. North and C.M. Macal. *Managing business complexity: Discovering strategic solutions with agent-based modeling and simulation*. Oxford University Press, 2007.

[79] H.V. Parunak, R. Savit, and R.L. Riolo. Agent-based modeling vs. equation-based modeling: A case study and user's guide. In J. Sichman, R. Conte, and N. Gilbert, editors, *Multi-Agent Systems and Agent-Based Simulation*, volume 1534 of *Lecture Notes in Computer Science*, pages 10–25. Springer Berlin Heidelberg, 1998.

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

[81] V. Rai and A.L. Beck. Public perceptions and information gaps in energy conservation and solar energy in texas. *Working Paper*, http://bit.ly/Energy-Perceptions-In-Texas, 2015.

[82] V. Rai and K. McAndrews. Decision-making and behavior change in residential adopters of solar PV. In *Proceedings of the World Renewable Energy Forum*, Denver, CO, May 2012.

[83] 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), 2013.

[84] V. Rai and B. Sigrin. Diffusion of environmentally-friendly technologies: Buy vs. lease decisions in residential PV markets. *Environmental Research Letters*, 8(1), 2013.

[85] S.F. Railsback and V. Grimm. *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton University Press, 2011.

[86] 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.

[87] 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.

[88] D.T. Robinson and D.G. Brown. Evaluating the effects of land-use development policies on ex-urban forest cover: An integrated agent-based gis approach. *International Journal of Geographical Information Science*, 23(9):1211–1232, 2009.

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

[90] A. Rubinstein. *Modeling Bounded Rationality*, volume 1. MIT press, 1998.

[91] R.K. Sawyer. *Social emergence: Societies as complex systems*. Cambridge University Press, 2005.

[92] S. Schnettler. A structured overview of 50 years of small-world research. *Social Networks*, 31(3):165–178, 2009.

[93] 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.

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

[95] X. Shi, K. Fernandes, and P. Chumnumpan. Diffusion of multi-generational high-technology products. *Technovation*, 34(3):162–176, 2014.

[96] A. Smajgl and O. Barreteau. *Empirical agent-based modeling: Challenges and solutions*. Springer, 2014.

[97] 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.

[98] 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.

[99] P.C. Stern. What psychology knows about energy conservation. *American Psychologist*, 47(10):1224, 1992.

[100] 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(2):261–272, 2008.

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

[102] V.V. Tyagi, N.A.A. Rahim, N.A. Rahim, and Jeyraj A.L. Selvaraj. Progress in solar PV technology: Research and achievement. *Renewable and Sustainable Energy Reviews*, 20(0):443 – 461, 2013.

[103] U.S. Energy Information Administration. Electric power annual. Technical report, U.S. Department of Energy, 2013.

[104] T.W. Valente. Social network thresholds in the diffusion of innovations. *Social Networks*, 18(1):69–89, 1996.

[105] A. Van Benthem, K. Gillingham, and J. Sweeney. Learning-by-doing and the optimal solar policy in California. *The Energy Journal*, pages 131–151, 2008.

[106] O. Van Vliet, B. Vries, A. Faaij, W. Turkenburg, and W. Jager. Multi-agent simulation of adoption of alternative fuels. *Transportation Research Part D: Transport and Environment*, 15(6):326–342, 2010.

[107] H. Visser and T. de Nijs. The map comparison kit. *Environmental Modelling & Software*, 21(3):346–358, 2006.

[108] F. Wang. Agent-based control for networked traffic management systems. *Intelligent Systems, IEEE*, 20(5):92–96, 2005.

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

[110] S.R. Wealands, R.B. Grayson, and J.P. Walker. Quantitative comparison of spatial fields for hydrological model assessment–some promising approaches. *Advances in Water Resources*, 28(1):15–32, 2005.

[111] C. Werker and T. Brenner. Empirical calibration of simulation models. Technical report, Papers on Economics and Evolution, 2004.

[112] C. Wilson and H. Dowlatabadi. Models of decision making and residential energy use. *Annual Review of Environment and Resources*, 32:169–203, 2007.

[113] P. Windrum, G. Fagiolo, and A. Moneta. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation*, 10(2):8, 2007.

[114] W. Wood. Attitude change: Persuasion and social influence. *Annual Review of Psychology*, 51(1):539–570, 2000.

[115] B. Zhang, Y. Zhang, and J. Bi. An adaptive agent-based modeling approach for analyzing the influence of transaction costs on emissions trading markets. *Environmental Modelling & Software*, 26(4):482–491, 2011.

[116] T. Zhang. Agent-based simulation of consumer purchase decision-making and the decoy effect. *Journal of Business Research*, 60(8):912–922, 2007.

[117] 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.