CrossMark

# Empirically grounded agent-based models of innovation diffusion: a critical review

**Haifeng Zhang**[1] · **Yevgeniy Vorobeychik**[1]

**Abstract** Innovation diffusion has been studied extensively in a variety of disciplines, including sociology, economics, marketing, ecology, and computer science. Traditional literature on innovation diffusion has been dominated by models of aggregate behavior and trends. However, the agent-based modeling (ABM) paradigm is gaining popularity as it captures agent heterogeneity and enables fine-grained modeling of interactions mediated by social and geographic networks. While most ABM work on innovation diffusion is theoretical, empirically grounded models are increasingly important, particularly in guiding policy decisions. We present a critical review of empirically grounded agent-based models of innovation diffusion, developing a categorization of this research based on types of agent models as well as applications. By connecting the modeling methodologies in the fields of information and innovation diffusion, we suggest that the maximum likelihood estimation framework widely used in the former is a promising paradigm for calibration of agent-based models for innovation diffusion. Although many advances have been made to standardize ABM methodology, we identify four major issues in model calibration and validation, and suggest potential solutions.

**Keywords** Literature review · Innovation diffusion · Agent-based modeling · Empirical method · Calibration · Validation

## 1 Introduction

### 1.1 Innovation diffusion: theoretical foundations

*Diffusion* refers to the process by which an innovation is adopted over time by members of a social system (Rogers 2010; Valente 2005). Commonly, an *innovation* refers to a new

---

✉ Haifeng Zhang
    haifeng.zhang@vanderbilt.edu

    Yevgeniy Vorobeychik
    yevgeniy.vorobeychik@vanderbilt.edu

[1] Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA

Springer

technology, but the conceptual notion can be applied far more broadly to consider the spread of ideas and practices. Rogers (2010) laid down the theoretical foundations of innovation diffusion in his book, *Diffusion of Innovations*, in which he synthesizes studies in anthropology, sociology, and education, and proposes a generic theory to explain the diffusion of innovations among individuals and organizations. He suggests five characteristics of innovation to determine the rate of adoption: *relative advantage*, *compatibility*, *complexity*, *trialability*, and *observability*. Rogers models human decision about adoption of innovation as a multi-stage process, involving five stages: *knowledge*, *persuasion*, *decision*, *implementation*, and *confirmation*. Furthermore, he classifies individuals into five adopter categories: *innovators*, *early adopters*, *early majority*, *late majority*, and *laggards*. In addition to these high-level considerations, much attention has been on the significance of social relationships and influence in innovation diffusion (in contrast with, or complementary to, economic considerations). Starting with early groundwork (Ryan and Gross 1943), there has now been extensive research on how social network structure, group norm, opinion leadership, weak ties, and critical mass impact diffusion of innovations (Valente 1995; Valente and Rogers 1995).

## 1.2 Mathematical models of innovation diffusion

Traditional mathematical models of innovation diffusion aim to model aggregate trends, rather than individual decisions. Numerous such models follow the framework of Bass model, which is one of the most influential models in marketing (Bass 1969; Hopp 2004). The Bass model was originally designed for forecasting sales of new consumer durables. The model assumes that the probability of adopting a product, given the person has not yet adopted, is linearly related to the number of past adopters. The Bass model can be calibrated with aggregate sales data, and Bass showed that it can qualitatively capture the S-shaped pattern of aggregate adoption over time (Peres et al. 2010).

The Bass model has a number of limitations. First, it does not capture individual interactions. Indeed, the model explicitly assumes a fully connected and homogeneous network. For innovation diffusion, this is an important drawback, as individual interdependence and communications are among the most significant aspects to understand innovation diffusion (Valente 2005; Rogers 2010). The second criticism of the Bass model is that it does not include any decision variables that are of interest from a managerial perspective. The issue has been addressed later by incorporating the marketing mix variables, price, and advertising, into the diffusion model. For an extensive review of research in this direction, we refer readers to Mahajan et al. (2000) and Meade and Islam (2006). Nevertheless, these marketing mix variables are mostly designated for the entire market without a consideration of individual heterogeneity. Lastly, the predictability of the Bass model is often questioned. For example, Chandrasekaran and Tellis (2007) argue, that the model needs considerable data around the critical point at which diffusion accelerates to be effective, but once such data is available the value of the Bass model becomes limited.

## 1.3 Agent-based modeling for innovation diffusion

Agent-based modeling (ABM) has emerged as another natural approach to study innovation diffusion. Agent-based models are typically simulation models that capture dynamic interactions among a (usually large) collection of individuals. They were originally developed as a tool for complexity theory research (Lewin 1999; Holland 1995), and have gained popularity in many scientific areas for the past decade (Gilbert and Troitzsch 2005; Macal and North 2010; Garcia and Jager 2011; Macal 2016). The ABM paradigm offers two advantages for the

study of innovation diffusion: first, it facilitates the modeling of agent heterogeneity, and second, it enables fine-grained modeling of interactions mediated by social networks. Indeed, agent-based modeling has been applied in study of innovation diffusion to aid intuition, theoretical exploration, and to provide policy decision support (Kiesling et al. 2012).

Traditional agent-based models are largely conceptual (Axelrod 1997; Epstein 1999). This use of ABMs as primarily conceptual tools is partly because they are commonly considered as ideal *learning tools* for scientists to understand a system under a variety of conditions by simulating the interactions among agents. As a consequence, the simplicity of agent rules is commonly a crucial consideration in the design of agent-based models. Such simplicity, however, has given rise to criticism of the ABM methodology as being "toy models" that do not reflect reality (Garcia and Jager 2011). Moreover, an increasingly important criticism is that if ABMs are used in any policy decision support, the predictive validity of the model becomes paramount, and models that are primarily conceptual may be inadequate for such tasks.

It is this increasing use of agent-based modeling to obtain policy guidance that has motivated increasing use of empirically grounded agent-based models. Empirical agent-based models have recently experienced significant growth (Kiesling et al. 2012). In these studies, empirical data are used to initialize simulation, parameterize agent-based models, or to evaluate model validity. The explosion of high-resolution data sets, coupled with advances in data analytics and machine learning have given rise to increased opportunities for empirically grounding agent-based models, and this trend is likely to continue. Our goal is to provide an overview of these empirically grounded agent-based models developed with the goal of studying innovation diffusion. Through a careful examination of these studies, we also aim to identify potential methodological issues that arise, and suggest ways to address these.

## 1.4 Contributions

The diffusion of new products has been an important topic for decades (Mahajan et al. 1990, 2000; Meade and Islam 2006; Chandrasekaran and Tellis 2007; Peres et al. 2010). The prevalence of ABM approach can be glimpsed from a number of review papers from disciplines like sociology (Macy and Willer 2002), ecology (Matthews et al. 2007), and marketing (Garcia 2005; Hauser et al. 2006; Negahban and Yilmaz 2014). For example, Garcia (2005) describes potential uses of ABM in market research associated with innovations, exploring benefits and challenges of modeling complex dynamical systems in this fashion. Dawid (2006) surveys agent-based models of innovation diffusion within a computational economics context. Peres et al. (2010) review diffusion models in the context of a single market and cross-markets and brands. To the best of our knowledge, the closest work to ours is a review of agent-based simulations of innovation diffusion by Kiesling et al. (2012), who survey both theoretical and empirical work. In comparison with these past reviews, we make the following novel contributions:

1. We provide *systematic* review of the empirical agent-based models of innovation diffusion. This is in contrast to the narrative review of the applied work as provided in Kiesling et al. (2012). In particular, we offer a novel classification of agent adoption models as employed in the reviewed papers. By highlighting the adoption models and their parameterization methods, we aim to bridge methodological gaps among domains and applications. We identified the papers to include in a rigorous and systematic manner. In terms of scope, any work presenting an agent-based model using empirical data to simulate the diffusion of innovations was included. Our selection process combined results from multiple databases, including Google Scholar and ScienceDirect, with extensive

search for relevant keywords, and back-tracking and forward-tracking reference lists, while carefully screening out non-candidates.

2. Our review is *comprehensive* and *updated*. The collection of reviewed papers spans a *superset* of the applications as covered in Kiesling et al. (2012) and, indeed, a number of significant efforts have emerged after 2012. Notably, we also include a selection of papers from the literature on information diffusion, a fast-growing area. These models rely on principled machine learning techniques for model calibration based on empirical observations of diffusion traces. In addition, we exclude two (out of 15) papers from Kiesling et al. (2012) which are not empirically grounded. In the end, we reviewed 43 papers, of which 30 (23 from years after 2011) were not included by Kiesling et al. (2012).

3. We provide a *critical* review, assessing strength and weaknesses of the surveyed research. Almost all surveyed papers followed standard modeling steps and presented their results systematically. However, we conclude that the current literature commonly exhibits several major shortcomings in model calibration and validation.[1] Addressing these issues would significantly increase the credibility of agent-based models. We, therefore, devote a section to an overview of existing validation methods in the literature and an in-depth discussion of these issues and potential solutions.

## 2 Categorization of empirically grounded ABMs of innovation diffusion

We review the burst of recent developments of empirically grounded agent-based models, which are examined through two dimensions: models and applications. First, to facilitate methodological comparison, we group the papers into six categories which represent the specific approaches taken to model individual agent decision processes: *mathematical optimization based models*, *economics-based models*, *cognitive agent models*, *heuristic models*, *statistics-based models* and *social influence models*. Second, as we observe that modeling efforts span several domains, the next section offers an application-focused categorization.

The categorization in this section is aimed at qualitatively clustering the existing agent-based models with respect to their *modeling methods*, which can be further characterized from several dimensions, such as behavioral assumption, data granularity, internal structure, calibration and validation. The six categories we identified present a comprehensive picture and structured patterns of the different methods used to model individual agent decision processes seen in a variety of applications.

We review each paper in sequence and in some detail, providing sufficient depth in the review for a reader to understand the nature of each surveyed work. In particular, we focus on how data was used in the modeling process, and in particular, in initialization, calibration and validation steps. We attempt to draw connections among the papers using our categorization structure (i.e., by grouping them into the six categories based on the methodology used to model individual agent behavior). Table 1 shows how these survey articles are distributed across the categories and publication years.[2] Notice that this approach is different from the synthesis-based approach followed by other review papers, such as, Windrum et al. (2007) and Macal (2016), which generally draws conclusion for a collection of papers but does not provide sufficient detail to assess how data is used in these efforts.

---

[1] The concepts of calibration and validation are explained in Sect. 5.1 below.

[2] For simplicity, we omit "20" and use the last two digits to denote a year. For example, "07(2)" stands for 2 publications in year 2007.

**Table 1** Distribution of surveyed papers over categories and years

| Category by modeling methods | Distribution in year | Total published |
| --- | --- | --- |
| Mathematical optimization based model | 01, 07(2), 09, 10, 13 | 6 |
| Economics-based model | 10, 11(2), 12, 13, 14(2), 15 | 8 |
| Cognitive agent model | 02, 06, 09(2), 12, 13(2), 15(2), 16(2) | 11 |
| Heuristic model | 10, 11(2) | 3 |
| Statistics-based model | 07(2), 08, 09, 11, 12, 13, 14, 15, 16 | 10 |
| Social influence model | 13(2), 14, 16, 17 | 5 |
| Total | | 43 |

## 2.1 Mathematical optimization (MO) based models

The MO-based models posit that agents (e.g., farmer households) are deliberate decision-makers who use sophisticated mathematical planning tools to assess the possible consequence of actions. While agents may encounter uncertainty, incomplete information, and constraints, their final decisions to adopt innovations are determined by concrete optimization objectives. The use of complex mathematical programs is commonly justified by the fact that farmer agents often consider their farming decisions in terms of economic returns.

In a seminal paper, Berger (2001) developed a spatial multi-agent mathematical programming (MP) model of diffusion of farming innovations in Chile. Production, consumption, investment, and marketing decisions of individual households are modeled using linear programming with the goal of maximizing expected family income subject to limited land and water assets. Moreover, in accordance with the literature on innovation diffusion, the model incorporates effects of past experience, as well as observed experience by peers. This is done by imposing a precondition for the MP procedure that the net benefit is only calculated if peer adoption level reaches the predefined threshold. In addition to such contagion effects, agent interactions are also reflected by the feedback effects of land and water resources and return-flows of irrigation water, implemented by coupling the economic agent decision model with hydrological components. In simulation models, agents are cellular automata with each cell associated with biophysical and economic attributes, such as soil quality, water supply, land cover/land use, ownership, internal transport costs, and marginal productivity. *These agent properties are initialized using empirical data* derived from various data sources, including a survey that captures both agronomic and socio-economic features, and a spatial data set with information about land and water use. *Parameters were calibrated* in terms of closeness of simulation experiments and farm data at both macro and micro levels. *Validation* was then performed by regressing land use results based on the model on actual land use in the data. Although values of the slope of this regression are reported for both macro and micro levels, validation is incomplete. For instance, micro-validation is only conducted for the year when the simulation starts due to data availability. Finally, the fact that validation was not conducted on data independent from calibration is another important weakness. Later, Berger et al. (2007) applied his MP-based agent-based modeling approach to study the complexity of water uses in Chile. Unfortunately, that work still had the same issue on validation.

Schreinemachers et al. (2007) adopted the MP-based approach to simulate soil fertility and poverty dynamics in Uganda, and analyze the impact on these of access to short-term credit and alternative technologies. At the heart of the model is a simulation of a farmer's decision process, crop yields, and soil fertility dynamics. The decision model is comprised

of three parts: (1) a set of possible decisions related to agriculture, such as growing crops, raising livestock, and selling and purchasing agricultural products; (2) a utility function that determines how much the decisions contribute to the farmer's objectives; and (3) links among decision variables represented by a set of equations. Following Berger (2001), a three-stage decision flow is defined that separates agent decisions into investment, production, and consumption. Moreover, the portion of the model capturing consumption includes econometrically-specified allocation of farm and non-farm income to saving, food, and other expenditures. Properties of the household agent, such as quantity and quality of land, labor, livestock, permanent crops, and knowledge of innovation, are sampled from empirical distributions based on limited samples. Additional features include models of animal and tree growth, technology diffusion, demographics, and price changes. In technology diffusion, peer influence is captured in the same manner as Berger (2001), but notably, each agent is assigned a threshold based on household survey data. The model was systematically validated in three steps: first, econometric models were validated for accuracy, then each component was validated independently, and finally the system as a whole. Similar to Berger (2001), validation used the same data as calibration.

Schreinemachers et al. (2009) studied the diffusion of greenhouse agriculture, using bell pepper in a watershed in the northern uplands of Thailand as a case study. The work largely follows the MP-MAS (mathematical programing-based multi-agent systems) approach due to Berger (2001). Notably, the author proposes calibrating the diffusion thresholds as described in Berger (2001) by using a binary adoption model (e.g., logistic regression), which is estimated from farmer survey data. To obtain threshold values for individuals, the author first computes adoption probability for each agent based on a set of observable independent variables, and then ranks these, dividing them into the five categories of innovators due to Rogers (1995). Validation was carried by checking the value of $R^2$ associated with a regression of observed land use on its predicted value. The proposed validation method suffers from the same limitation as other related research in using the same data for calibration and validation.

Schreinemachers et al. (2010) applied the MP-based approach to study the impact of several agricultural innovations on increasing profitability of litchi orchards in Northern Thailand. Unlike Schreinemachers et al. (2009) that estimated a logistic regression model to assign agents to threshold groups, they assigned thresholds randomly due to the lack of relevant data. The model was validated using regression method as described in Schreinemachers et al. (2009), and validation suggests that the model reasonably represents aggregate agent behavior, even while individual-level behavior is not well captured. As in prior work, calibration and validation used the same data.

Alexander et al. (2013) developed an agent-based model of the UK perennial energy crop market to analyze spatial and temporal dynamics of energy crop adoption. The model includes the interaction of supply and demand between two agent groups: farmers and biomass power plant investors. The farmer agents have fixed spatial locations which determine the land quality and climate that in turn impact crop yields, and decide on the selection of crops via a two-stage approach similar to Berger (2001), with peer influence again modeled through a threshold function. A farmer agent considers adoption only if the proportion of neighbors within a given radius with a positive adoption experience exceeds a threshold. When adoption is considered, a farm scale mathematical program is used to determine the optimal selection of crops that maximizes utility as described in Alexander et al. (2014). Calibration of the farm scale model is either informed by empirical data or in reference to previous studies. Validation involved checking model behaviors on simplified configurations,

unit-testing of model components, and comparing simulation results against empirical data. However, validation did not use independent data from calibration.

## 2.2 Economics-based models

Unlike the MO based models in Sect. 2.1, the *economics*-based models use simpler rules with fewer constraints and decision variables. Particularly, agents commonly simply minimize cost, maximize profit, or, more generally, maximize personal utility.

### 2.2.1 Cost minimization

Faber et al. (2010) develop an agent-based simulation model for energy technologies, micro-CHP (combined heat and power) and incumbent condensing boilers, in competition for consumer demand. Consumer agents are classified by housing type, which is viewed as the most important factor in determining natural gas requirements for heating units. At each time step a consumer considers purchasing a new heating unit, and follows a three-step decision algorithm: (1) assess if a new unit is needed, (2) scan the market for "visible" heating units, where "technology awareness" is formulated as a function of the level of advertising, market share, and bandwagon effect, and (3) each consumer chooses the cheapest technology of those that are visible. The cost, which depends on the consumer's class, is comprised of purchase costs, subsidies, and use costs over the expected life of the technology. Some of the parameters are calibrated using empirical data, while others are set in an ad hoc fashion. Some validation was performed through the use of a sensitivity analysis of the variables such as market size, progress rate, and technology lifetime. However, no explicit model validation using empirical data was undertaken.

### 2.2.2 Profit maximization

Sorda et al. (2013) develop an agent-based simulation model to investigate electricity generation from combined heat and power (CHP) biogas plants in Germany. Instead of simulating farmer's individual decision whether to invest in a biogas plant, the model solves a system-wide optimization problem from the perspective of a global planner. The model includes two types of agents: information agents, including Federal Government, Bank, Electric Utility, and Plant Manufacturer, and agents making investment decisions, including the Substrate Supplier, District, Decision-Maker, and Heat Consumer. The core decision-making agent acts as a representative for investors in each community. The agent chooses to invest in a biogas facility whenever sufficient resources are available and the investment yields positive net present value. This work used multiple data sources to construct the simulation model. For example, plant operator guidelines and manufacturer specifications were used to obtain data about the characteristics of biogas plants. Although the model is thus informed by real data, it is not quantitatively validated.

### 2.2.3 Utility maximization

Broekhuizen et al. (2011) develop an agent-based model of movie goer behavior which incorporates social influence in movie selection decisions. Their study investigates two types of social influence: the influence of past behavior by others, and influence stemming from preferences of an individual's friends, such as group pressure to join others in seeing a

movie. The main purpose of this work is to determine the degree to which different types of social influence impact inequality. In their model, agent's decision-making is probabilistic and utility-driven. An agent first observes which movies are being shown in the marketplace with some probability. Next, with a specified probability, an agent is selected to consider seeing a movie. If selected, it goes to the movie that maximizes expected utility among all those it is aware of. Otherwise, it does not see any movie. *Utility* in this setting is a weighted sum of *individual* utility, which represents the alignment between individual's preferences and movie characteristics, and *social* utility which is a combination of the two types of social influence above. Some of the model parameters are either theoretically determined or empirically calibrated, while the variability of the rest is investigated by sensitivity analysis. Validation involved a cross-national survey, using cross-cultural differences due to Hofstede's collectivism-individualism index to measure social influence. While the validation is based on an independent survey study, it is largely qualitative.

Günther et al. (2011) introduce an agent-based simulation approach to support marketing activities. The proposed model was applied to the study of a new biomass-based fuel that would likely be introduced in Austria. Consumer agents are embedded in a social network, where nodes represent agents and edge weights determine the probability with which the connected agents communicate. The authors tested several network structures, including random (Erdos-Renyi) networks, small-world networks, and so-called "preferences-based" networks, where connections between agents are based on geographical and cognitive proximity as well as opinion leadership. Each agent is characterized by preferences, geographical position, tanking behavior, how informed they are about the product, and their level of social influence. Agents have preferences for several product attributes: price, quality, and expected environmental friendliness, which are initialized differently based on consumer type. Agents are geographically distributed in virtual space based on the spatial distribution of Austrian population, and their tanking behavior is a function of fuel tank capacity, travel behavior, and habits. Individual information level on the innovation at hand captures the knowledge about a product, which increases as a function of interpersonal communication and exposure to marketing activities. Influence level, on the other hand, represents an agent's expertise with the innovation and determines the amount of information received through communication. Upon interaction, an agent with lower information level learns from a more informed agent. Most importantly, the *utility* function for agent $i$ at time $t$ is given by $u_{i,t} = (1 - Price_t) \times w_{i,1} + Price_t \times w_{i,2} + ppq_{i,t} \times w_{i,3} + w_{i,4}$, where $0 \leq w_{i,k} \leq 1$ and $\sum_{k=1}^{4} w_{i,k} = 1$, and the first and second weights pertain to price, while the last two represent how strongly agents prefer quality and how willing they are to seek renewable energy sources for fuel, respectively. An agent is assumed to adopt if utility exceeds a specified individual threshold drawn for each agent from the uniform distribution. Moreover, the *perceived product quality*, $ppq_i$ is assumed to gradually converge the true product quality for adopters. The author briefly mentions that model parameters are set in reference to a prior case study. Apart from this, no detailed information is provided about how model parameters are actually calibrated in the setting. Moreover, the model was only validated qualitatively with subjective expert knowledge.

Holtz and Pahl-Wostl (2012) develop a utility-based agent-based model to study how farmer characteristics affect land-use changes in a region of Spain. As relevant data are scarce, their model cannot be quantitatively calibrated and validated. Empirical data are used to initialize the model, deriving the initial crop distribution, and to assess the validity of the model qualitatively. In this model, an agent's utility is formulated as a Cobb–Douglas function by multiplying four influences: gross margin, risk, labor load, and regulatory constraints.

Parameters associated with these influences differ with the types of farmers, for example, part-time, family, and business-oriented farmers would have distinct utility parameters. In the decision process, an agent chooses a land use pattern that maximizes its utility, where land use patterns involve a combination of crop and irrigation technology, constrained by policies. The diffusion of irrigation technology is simulated based on the concept that the more widely used a technology is, the more likely it is to be considered by individual farmers. Their experiments explore the importance of each influence variable in the utility function, as well as of farmer types, by qualitatively comparing the simulation results with empirical data.

Plötz et al. (2014) propose a model for the diffusion of electric vehicles (EVs) to evaluate EV-related policies based driving data in Germany. The model determines the market shares of different technologies by simulating each driving profile as both EV and conventional vehicle, choosing the option which maximizes the driver's utility, and then extrapolating these agent-level choices to aggregate market shares. In modeling individual decisions, utility is defined as a function of *total cost of ownership* (TCO), choice of EV brands, and individual *willingness-to-pay-more* (WTPM). The authors combined survey results with driving profiles to derive four categories of agents (adopters), and assigned each driving profiles to one of these categories. Through simulating the *plug-in hybrid electric vehicle* (PHEV) share of the market as a function of annual average *vehicle kilometers traveled* (VKT) for medium-sized vehicles, the model was validated by comparing original group assignment with simulated outcomes and by examining simulated diesel market shares relative to actual values within different branches of industry. While validation is quantitative and rigorous, it does not use independent data. Moreover, the model does not capture social influence which is often a key aspect of innovation diffusion modeling.

McCoy and Lyons (2014) develop an agent-based model of diffusion of electric vehicles among Irish households. Agents representing households are located at a regular lattice space. They are heterogeneous as suggested by their characteristics. Agents have two static attributes, *Income Utility* (IU) and *Environmental Utility* (EU), drawn independently from empirical distributions derived from a survey. In particular, IU is based on an agents social class, tenure type, and age, which are assumed to be highly correlated with income, whereas EU is based on the agent's past adoption of energy efficiency technologies and their attitude toward the environment. Each agent $i$ has a unique threshold, $\theta_i$, drawn from a distribution that is negatively correlated to IU, and adopts if $U_i(t) \geq \theta_i$ and $t \times crit \geq rand(0, 1)$, where, $crit$ is decimal value that is used to account for inertia that exists in early stage of technology adoption, while utility $U_i(t)$ is defined as $U_i(t) = \alpha_i IU_i + \beta_i EU_i + \gamma_i G_i(t) + \delta_i S(t)$, where, $IU$ represents individual's preferences, $G$ is social influence, and $S$ is social norms, and $\alpha_i + \beta_i + \gamma_i + \delta_i = 1$. To allow these parameters to vary by agent, the authors specify four distinct consumer groups with different preferential weighting schemes. Although the agents in the simulation are initialized using empirical distributions, key parameters in the decision model are not derived empirically but are based on the authors' assumptions. Additionally, no rigorous validation is provided.

Palmer et al. (2015) developed an agent-based model of diffusion of solar photovoltaic (PV) systems in the residential sector in Italy. The *utility* of agent $j$ is defined as the sum of four weighted partial utilities, i.e., $U(j) = w_{pp}(sm_j) \cdot u_{pp}(j) + w_{env}(sm_j) \cdot u_{env}(j) + w_{inc}(sm_j) \cdot u_{inc}(j) + w_{com}(sm_j) \cdot u_{com}(j)$, where $\sum_k w_k(sm_j) = 1$ for $k \in K : \{pp, env, inc, com\}$ and $w_k(sm_j), U(j) \in [0, 1]$. From left to right the partial utilities are: (1) payback period of the investment, (2) environmental benefits, (3) household income, and (4) social influence. An agent chooses to invest in PV if its total utility exceeds an exogenously specified threshold. Thresholds above vary by agent's demographic and behavioral characteristics, $sm_j$. The four partial utilities are derived from empirical data.

Specifically, the payback period is estimated based on investment costs, local irradiation levels, government subsidies, net earnings from generating electricity from the system vs. buying it from the grid, administrative fees, and maintenance costs. The environmental benefit is based on an estimate of reduced $CO_2$ emissions saved. Household income is estimated based on household demographics, such as age, level of education, and household type. Finally, social influence is captured by the number of neighbors of a household within its social network who have previously adopted PV. The social network among agents is generated according to the small-world model (Watts and Strogatz 1998), modified to account for socio-economic factors. The model parameters are calibrated by trying to match simulated adoption with the actual aggregate residential PV adoption in Italy over the 2006–2011 period. The model is then applied to study solar PV diffusion in Italy over the 2012–2026 period. However, no quantitative validation is offered.

## 2.3 Cognitive agent models

While both MO-based (Sect. 2.1) and economics-based (Sect. 2.2) models elaborate economic aspects of the decision process and integrate simple threshold effects, cognitive agent models aim to explicitly model how individuals affect one another in cognitive and psychological terms, such as opinion, attitude, subjective norm, and emotion. This category includes the *Relative Agreement Model*, the *Theory of Planned Behavior*, the *Theory of Emotional Coherence*, and the *Consumat Model*.

### 2.3.1 Relative agreement model

The *Relative Agreement Model* belongs to a class of opinion dynamics models (Hegselmann et al. 2002) and addresses how opinion and uncertainty are affected by interpersonal interactions. Seminal work is due to Deffuant et al. (2000), who investigate how the magnitude of thresholds, with respect to attitude difference, leads to group opinion convergence and extremeness. The relative agreement model is often known as "*Deffuant model*" in the literature.

Deffuant et al. (2002b) design an agent-based model to simulate organic farming conversion in France. To model impact of interactions on the individual decision, they relied on the Deffuant model in which both opinion and uncertainty are continuous variables. In the diffusion model, farmer agent has an "interest" state with three possible values: *not-interested*, *uncertain*, and *interested*. The actual value is based on the agent's opinion (represented as a mean value and confidence) and economic consideration. The value of the interest state depends on the position of the global opinion segment compared to a threshold value. Agent changes opinion after discussing with peers using a variant of the *Relative Agreement* algorithm (Deffuant et al. 2002a). The farmers send messages containing their opinions and information, following a two-stage diffusion model of Valente (1995), mediated by a network generated according to the Watts and Strogatz (1998) model. These impact opinions of the recipients as a function of opinion similarity, as well as confidence of the sender, with more confident opinions having greater influence. In addition, if the farmer agent is "interested" or "uncertain", he performs an evaluation of the economic criterion, and if he remains interested, he requests a visit from a technician. After this visit, the economic criterion is evaluated again under reduced uncertainty. Finally, the adoption decision is made when the farmer has been visited by a technician and remains "interested" for a given duration.

Many model parameters governing the decision and communication process are not informed by empirical data. The authors tested the sensitivity of the model by varying these

variables, including the main parameters of the dynamics, the parameters of the initial opinion distribution average number of neighborhood and professional links, and variations of the institutional scenario. Within this parametric space, they aimed to identify parameter zones that are compatible with empirical data. For each parameter configuration, the authors defined two error measures: the adoption error and the error of proximity of adopters to the initial organic farmers. A decision tree algorithm was then used to find the parameter zones where the simulated diffusion has an acceptable performance. While this sensitivity analysis step can be viewed as model calibration, it is distinct from classical calibration which aims at finding a single best parameter configuration. The model was not validated using independent data.

### 2.3.2 Theory of planned behavior

The *Theory of Planned Behavior (TPB)* postulates that an individual's *intention* about a behavior is an important predictor of whether they will engage in this behavior (Ajzen 1991). As a result, the theory identifies three attributes that jointly determine intention: *attitudes*, *subjective norms*, and *perceived behavioral control*. The relative contribution for each predictor is represented by a weight which is often derived empirically using regression analysis based on survey data.

Kaufmann et al. (2009) build an agent-based simulation model on TPB to study the diffusion of organic farming practices in two New European Union Member States. Following the TPB methodology, each agent is characterized by three attributes: the attitude $a_i$, subjective norm $s_i$, and perceived behavioral control $p_i$, each ranging from -1 (extremely negative) to +1 (extremely positive). The intention $I_i$ is defined as $I_i = w_i^a a_i + w_i^s s_i + w_i^p p_i$, where $w_i^a, w_i^s, w_i^p$ are relative contribution toward intention. The weights for non-adopters and adopters are derived separately using linear regressions based on the survey data. If an agent's intention exceeds a threshold $t$ it adopts, and does not adopt otherwise. The threshold is obtained from survey data as the average intention of non-adopters who have expressed a desire to adopt. In the simulation model, social influence is transmitted among network neighbors in each time step in a random order. Specifically, when one node speaks to another, the receiver shifts its subjective norm closer to the sender's intention, following the relative agreement framework. Social networks are generated to reflect small-world properties (Watts and Strogatz 1998) and a left-skewed degree distribution (Noble et al. 2004), with specifics determined by a set of parameters, which are set based on survey data (such as the average degree). While empirical data is thus used to calibrate parameters of the model, no quantitative validation was provided.

Schwarz and Ernst (2009) propose an agent-based model of diffusion of water-saving innovations, and applied the model to a geographic area in Germany. Agents are households with certain lifestyles, represented by demographic and behavioral characteristics. They use two different decision rules to determine adoption: a cognitively demanding decision rule representing a *deliberate* decision and a simple decision *heuristic*. The particular decision rule to use is selected based on the agent's type and technology category. The deliberate decision-making algorithm is based on multi-attribute subjective utility maximization that integrates attitude, social norm, and perceived behavioral control. The heuristic decision rule makes decisions in greedy order of evaluation criteria based on innovation characteristics and social norms. Finally, if no clear decision can be made, agents imitate their peers, who are defined through a variation of a small-world network (Watts and Strogatz 1998) which captures spatial proximity and lifestyle affinity in determining links among agents. The model was calibrated using data from a survey according to the framework of the Theory

of Planned Behavior (Ajzen 1991), with the importance of different decision factors derived by structural equation models or linear regressions for lifestyle groups. The model was validated using *independent* market research data at the household level. In addition, due to the lack of independent aggregated diffusion data, results of the empirical survey were used for validation.

Sopha et al. (2013) present an agent-based model for simulating heating system adoption in Norway. Their model extends TPB to consider several contributing factors, such as household groups, intention, attitudes, perceived behavioral control, norms, and perceived heating system attributes. Households are grouped using cluster analysis based on income level and basic values available in the survey data to approximate the influence of lifestyle on attitudes towards a technology. Attribute parameters are then estimated using regressions for each household cluster based on the household survey. Moreover, motived by the meta-theory of consumer behavior (Jager 2000), the model assumes that a household agent randomly follows one of four decision strategies: repetition, deliberation, imitation, and social comparison, in accordance with empirical distribution based on survey data. Notably, this model is validated using *independent* data that is not used for calibration, examining how well simulation reproduces actual system behavior at both macro and micro level.

Rai and Robinson (2015) develop an empirically grounded agent-based model of residential solar photovoltaic (PV) diffusion to study the design of PV rebate programs. The model is motived by TPB and assumes that two key elements determine adoption decision: attitude and (perceived) control. The authors calibrate population-wide agent attitudes using survey data and spatial regression. Following the opinion dynamics model in Deffuant et al. (2002a), at each time-step, agents' attitudes about the technology and their uncertainties are adjusted through interactions with their social network neighbors following the relative agreement protocol. Social influence is captured by households situated in small-world networks, with most connections governed by geographic and demographic proximity. In the "control" module, an agent $i$ compares its perceived behavioral control $pbc_i$ with the observed payback at the current time period $PP_{it}$. Then, if the agent exceeds its attitude threshold, it adopts when $PP_{it} < pbc_i$. $pbc_i$ for each agent $i$, is calculated as a linear sum of financial resources, the amount of sunlight received, and the amount of roof that is shaded, while $PP_{it}$ is calculated based on electricity expenses offset through the use of the solar system, the price of the system, utility rebates, federal investment tax credit, and annual system electricity generation. The six model parameters used to specify the social network, opinion convergence, the distribution of the behavioral control variable, and the global attitude threshold value were calibrated by an iterative fitting procedure using historical adoption data. The model was first validated in terms of predictive accuracy, comparing predicted adoption with empirical adoption level for the time period starting after the last date for the calibration dataset. Moreover, temporal, spatial, and demographic validation were conducted. However, validation was focused on aggregate (macro), rather than individual (micro) behavior.

Jensen et al. (2016) develop an agent-based model to assess energy-efficiency impacts of an air-quality feedback device in a German city. A household agent makes two decisions: whether to adopt a feedback device and whether to practice a specific energy-saving behavior. The model involves simulating both the adoption of the feedback device and the heating behavior respectively. Two diffusion processes are connected based on the observation that the feedback device changes an agent's heating behavior, and eventually will form a habit. In the simulations, household agents are generated based on marketing data on lifestyle, and initial adopters of the heating behavior are selected based on a survey. The adoption of an energy-efficient heating behavior is triggered by external events, whose rate is estimated by historical data using Google search queries. Their survey reveals that both information and

social influence drive behavior adoption. This insight is integrated into a decision-making model following the theory of planned behavior (TPB), in which information impacts the agent's attitude in each simulation step. On the other hand, the diffusion model of the feedback device is an adaptation of an earlier model also based on TPB. An adopter of the device is assumed to adopt the desired heating behavior with a fixed probability, which is informed by an empirical study. The space of model parameters is reduced by applying a strategy called "pattern-oriented modeling", which refines the model by matching simulation runs with multiple patterns observed from empirical data (Grimm et al. 2005). In their experiments, the authors calibrated several different models using empirical data and aimed to quantify the effect of feedback devices by comparing results generated by these models. However, no rigorous model validation is presented.

### 2.3.3 Theory of emotional coherence

When it comes to explaining and predicting human decisions in a social context, some computational psychology models also take emotional factors into account, which are often neglected by TPB-based models. Wolf et al. (2012) propose an agent-based model of adoption of electric vehicles by consumers in Berlin, Germany, based on the *Theory of Emotional Coherence (TEC)*. The parameters of the model were derived based on empirical data from focus groups and a representative survey of Berlin's population. In particular, the focus group provided a detailed picture of people's needs and goals regarding transportation; the survey was designed to generate quantitative estimates of the beliefs and emotions people associate with specific means of transportation. The attributes of the agents include age, gender, income, education, residential location, lifestyle, and a so-called social radius, and are obtained based on the survey data. The social network structure is generated by similarities between these characteristics following the theory of *homophily* (McPherson et al. 2001); specifically, the likelihood of two individuals communicating with one other is a function of their similarity in terms of demographic factors. To validate the predictions made by the model, the authors regressed empirical data related to actual transportation-related decisions (e.g., weekly car usage) from the survey on the activation parameters resulting from simulations. However, validation did not use independent data.

### 2.3.4 Consumat model

The Consumat Model is a social psychological framework, in which consumer agents switch among several cognitive strategies—commonly, *comparison*, *repetition*, *imitation*, and *deliberation*—as determined by need satisfaction and their degree of uncertainty (Jager et al. 2000). Schwoon (2006) uses an agent-based model (ABM) to simulate possible diffusion paths of fuel cell vehicles (FCVs), capturing complex dynamics among consumers, car producers, and filling station owners. In their model, the producers offer heterogeneous but similar cars, deciding in each period whether to change production to FCVs. Consumers have varying preferences for car attributes, refueling needs, and social influence factors. Although in a typical consumat approach (Janssen and Jager 2002), consumers follow one of four cognitive strategies on the basis of their level of need satisfaction and uncertainty, the author rules out repetition and imitation and argues that need satisfaction is rather low in their case. The consumer agent is assumed to maximize total expected utility, which is expressed as a function of car price, tax, the closeness between preferences and car characteristics, social need, as determined by the fraction of neighbors adopting each product type, and availability of

hydrogen. In the model, individual preferences may evolve with time to be more congruent with the "average car", as determined by a weighted average of attributes of cars sold in the previous period, where weights correspond to market shares. The model is calibrated by trying to match main features of the German auto market. The network structure governing social influence is assumed to form a torus. The model does not attempt quantitative validation.

### 2.3.5 The LARA model

LARA is the short for *Lightweight Architecture for boundedly Rational Agents*, a simplified cognitive agent architecture designed for large-scale policy simulations (Briegel et al. 2012). Comparing with existing complex psychological agent frameworks, LARA is more generalizable and easier to implement. We review two recent efforts motivated by the LARA architecture and grounded in empirical data.

Krebs et al. (2013) develop an agent-based model to simulate individual's provision of neighborhood support in climate change adaptation. In their model, agents are assigned to lifestyle groups and initialized using spatial and societal data. Motivated by LARA, an agent makes a decision in one of three modes: deliberation, habits, and exploration. In deliberation, an agent compares and ranks available options in terms of utility, which is the weighted sum of four goals: striving for effective neighborhood support, being egoistic, being altruistic, and achieving social conformity. The goal weights, which are different among lifestyle groups, are set based on expert ratings and the authors' prior work. A probability choice model is used to choose the final option when multiple better options are available. An agent acts in deliberation mode if no experience is available (habitual behavior is not possible) and shifts to an exploratory mode with a predefined small probability. The network in which the agents are embedded is generated using lifestyle information. Simulation runs for an initial period from 2001 to 2010 provide plausible results on behavioral patterns in cases of weather changes. From 2011 to 2020, the authors examine the effects of two intervention strategies that mobilize individuals to provide neighborhood support. Some model parameters remain uncalibrated, and the entire model is not validated due to a lack of empirical data at the macro level.

Krebs and Ernst (2015) develop an agent-based spatial simulation of adoption of green electricity in Germany. Each agent represents a household deciding to select between "green" and "gray" energy providers. Every agent is characterized by its geographical location and lifestyle group. Agents are initialized and parameterized by empirical data from surveys, psychological experiments, and other publicly available data. Following LARA, agents are assumed to make decisions either in a deliberative or habitual mode. Default agent behavior is habitual, and the agent transitions to a deliberative mode when triggered by internal and external events, such as a price change, personal communication, cognitive dissonance, need for cognition, and media events. An agent chooses an action that maximizes utility, which is a weighted sum of four goals: ecological orientation, economic orientation, social conformity, and reliability of provision. The goal weights depend on the lifestyle group and are derived from a survey and expert rating (Ernst and Briegel 2017). An artificial network that connects the agents is generated based on lifestyle and physical distance (Ernst and Briegel 2017). Once an agent decides to adopt green electricity, it chooses a service brand that is already known. The diffusion of the awareness of the brand is characterized by a simple word-of-mouth process. Validation focuses on two state variables of agent behavior: selected electricity provider and awareness of the brand, which involves comparing simulation results with

historical data both temporally and spatially starting from aggregate to the individual level. Unfortunately, validation was not conducted using independent data.

## 2.4 Heuristic models

Heuristic adoption models are often used when modelers are not aware of any established theories for agent decision-making in the studied application. These models tend to give us an impression of being "ad-hoc", since they are not built on any grounded theories. More importantly, unlike the cognitive agent models such as the theory of planned behavior, there is no established or principled means to estimate model parameters. Therefore, model parameters are often selected in order to match simulated output against a realistic adoption level. Although heuristic-based model appears to be an inaccurate representation of agent decision-making, they are easy to implement and interpret.

Van Vliet et al. (2010) make use of a *take-the-best* heuristic to model a fuel transportation system to investigate behavior of fuel producers and motorists in the context of diffusion of alternative fuels. In the model, producers' plant investment decision is determined by simple rules, and the same plant can produce multiple fuel types. Motorists are divided into several subgroups, each having distinct preferences. Each motorist is assumed to choose a single fuel type in a given year. Each fuel is assigned four attributes: driving cost, environment, performance, and reputation. Motorist preferences in the model are represented by two factors: (1) *priorities*, or the order of perceived importance of fuel attributes, and (2) *tolerance* levels, which determine how much worse a particular attribute of the corresponding fuel can be compared to the best available alternative to maintain this fuel type under consideration. The decision heuristic then successively removes the worst fuel one at a time in the order of attribute priorities. Due to the difficulty of obtaining actual preferences of motorists, the authors used the Dutch consumer value dispositions from another published model in literature as a proxy to parameterize the model. However, the model was not rigorously calibrated or validated using empirical data.

Zhao et al. (2011) propose a two-level agent-based simulation modeling framework to analyze the effectiveness of policies such as subsidies and regulation in promoting solar photovoltaic (PV) adoption. The lower-level model calculates payback period based on PV system electricity generation and household consumption, subsidies, PV module price, and electricity price. The higher-level model determines adoption choices as determined by attributes which include payback period, household income, social influence, and advertising. A pivotal aspect of the model is the *desire* for the technology (PV), which is formulated as a linear function of these four factors, and an agent adopts if the desire exceeds a specified threshold. Survey results from a prior study were used to derive a distribution for each factor, as well as the membership function in a fuzzy set formulation. The agents in the model were initialized using demographic data, along with realistic population growth dynamics based on census data. Moreover, calibration of threshold value was conducted to match simulated annual rate of PV adoption with historical data. However, the model was not quantitatively validated using independent data.

A more complex TOPSIS (*Technique for Order Preference by Similarity to Ideal Solution*) model is a decision heuristic which selects an option from several alternatives that is the closest to the ideal option and the farthest from the worst possible option. Kim et al. (2011) present agent-based automobile diffusion model using a TOPSIS approach to simulate market dynamics upon introduction of a new car in the market. The model integrates three determinants of purchasing behavior: (1) information offered by mass media, (2) relative importance of attributes to consumers, and (3) social influence. Individual agents rank

products by considering multiple product attributes and choosing a product closest to an ideal. A survey was conducted to estimate consumers' weights on the car attributes and the impact of social influence. In the simulations, diffusion begins with innovators who try out new products before others; once they adopt, their social network neighbors become aware of these decisions, with some deciding to adopt, and so on. A *small-world* network structure was assumed for this virtual market, and choices of rewiring and connectivity were determined by the model calibration step through comparing simulated results with historical monthly sales volumes of three car models. However, the model was not validated using independent data.

## 2.5 Statistics-based models

Statistics-based models rely on statistical methods to infer relative contribution of observable features towards one's decision whether to adopt. The estimated model is then integrated into an ABM. We review three subcategories of statistics-based methods for agent-based models of innovation diffusion: *conjoint analysis*, *discrete choice models*, and *machine learning*.

### 2.5.1 Conjoint analysis

Conjoint analysis is a statistical technique used in market research to determine how much each attribute of a product contributes to consumer's overall preference. This contribution is called the *partworth* of the attribute. Combining with feature values of innovation obtained from field study, one can construct a utility function accordingly.

Garcia et al. (2007) utilize conjoint analysis to instantiate and calibrate an agent-based marketing model using a case study of diffusion of Stelvin wine bottle screw caps in New Zealand. With a particular emphasis on validation, the overall work follows Carley (1996)'s four validation steps: *grounding*, *calibration*, *verification*, and *harmonizing* (the latter not performed, but listed as future work) to properly evaluate the model at both micro and macro levels. The model includes two agent types: wineries and consumers. In each period the wineries set the price, production level, and attributes of screw caps as a function of consumer demand. Consumers, in turn, make purchase decisions following their preferences. The model is calibrated using conjoint analysis, inferring *partworths* which determine consumer preferences in the model. Aggregate stylized facts were then replicated in the verification step. The work emphasizes the value of calibration, but pays less attention to validation, which is merely performed at a face level rather than quantitatively.

Vag (2007) presents a *dynamic* conjoint method that enables forecasts of future product preferences. The consumer behavior model considers many factors, including social influence, communication, and economic motivations. The author surveys behavior of individuals, such as their communication habits, and uses conjoint analysis to initialize preferences in the ABM. Notably, in this model agent priorities depend on one another, and the resulting social influence interactions may lead to large-scale aggregate shifts in individual priorities. To demonstrate the usability of their model, the study utilized empirical data on product preferences (in this case, preferences for mobile phones), consumer habits, and communication characteristics in a city in Hungary. Calibration of this model was only based on expert opinion and comparative analysis, rather than quantitative comparison with real data, and no quantitative validation was performed.

Zhang et al. (2011) develop an agent-based model to study the diffusion of eco-innovations, which in their context are alternative fuel vehicles (AFVs). The model considers interdependence among the manufacturers, consumers, and governmental agencies in the automotive

industry. The agents representing manufacturers choose engine type, fuel economy, vehicle type, and price, following a simulated annealing algorithm, to maximize profit in a competitive environment until a Nash equilibrium is reached (Michalek et al. 2004). The consumer agents choose which products to purchase. The partworth information in the utility function was derived by *choice-based conjoint* analysis using an empirical survey from Garcia et al. (2007). In particular, the probability of a consumer choosing a vehicle is formulated as a logit function of vehicle attributes, word-of-mouth, and domain-specific knowledge. The utility is modeled as a weighted sum of attributes, and parameters/partworth are estimated using hierarchical Bayes methods. The agent acting as "government" chooses policies aimed at influencing the behavior of both manufacturers and consumers. Model calibration involved conjoint analysis. However, the authors found that the ABM tended to overestimate the market shares of alternative fuel vehicles, which motivated them to adjust model parameters and to linearize the price parthworth in order to ensure that aggregate demand decreases with the price. Like Garcia et al. (2007), the authors follow the four steps of validation (Carley 1996). However, validation does not use data independent from calibration.

Lee et al. (2014) introduce an agent-based model of energy consumption by individual homeowners to analyze energy policies in the U.K. The model utilizes historical survey data and choice-based conjoint analysis to estimate the weight of a hypothetical utility function, defined as the weighted sum of attributes. In the simulation, moving and boiler break-down events are assumed to trigger a decision by the household agent. In this case, a particular alternative is selected if its utility is higher than all other alternatives as well as the status quo option. The model was populated with initial data based on a survey conducted in the U.K., and each agent was matched to a household type which can be further mapped to energy demand using energy consumption estimates. The authors then combined energy demand with fuel carbon intensity to determine annual household emissions. The model was calibrated by adjusting the weights in the decision model to match historic installation rates from 1996 to 2008 for loft insulation and cavity wall insulation. The model was not validated using independent data.

Stummer et al. (2015) devise an agent-based model to study the diffusion of multiple products. Each product is characterized by a number of attributes determined by expert focus group discussion. True performance of each product attribute is unknown to consumers, and each agent, therefore, keeps track of the distribution of attribute values based on information previously received. This information is updated based on interactions with peers, advertising, or direct experience. Consumer agent behavior is governed by a set of parameters that capture heterogeneous preferences and mobility behavior. Agents have additive multi-attribute utilities, the weights of which were obtained from survey data using conjoint analysis. The authors adapt the preferential attachment algorithm introduced by Barabási et al. (1999) to generate networks in which the attachment probability depends on both node degree and geographic distance between nodes. Network parameters were determined by taking into account additional information revealed in the consumer survey, such as the number of social contacts and communication frequency. An agent decides to purchase a product which maximizes utility. The model defines each advertising event to communicate a set of product attributes, which either increase product awareness or impact customer preferences. The model was validated extensively following Knepell and Arangno (1993), including conceptual validity, internal validity, micro-level external validity, macro-level external validity, and cross-model validity. The weakness of validation, however, is that it is only performed as an in-sample exercise without using independent data.

### 2.5.2 Discrete choice models

The *discrete choice* modeling framework, which originates in econometrics, is used to describe, explain, and predict agent choices between two or more discrete alternatives (Train 2009). The approach has a wide range of applications, and we review several efforts targeted specifically at innovation diffusion.

Galán et al. (2009) design an agent-based model to analyze water demand in a metropolitan area. This model is an integration of several sub-models, including models of urban dynamics, water consumption, and technological and opinion diffusion. The opinion diffusion model assumes that an agent's attitude towards the environment determines its water consumption, i.e., an non-environmentalist would use more water than an environmentalist. Accordingly, it is assumed that each agent can be in two states: *environmentalist* (E) or *non-environmentalist* (NE). The choice of a state depends on the agent's current state, the relative proportion of E and NE neighbors, and an exogenous term measuring the pressure towards E behavior. Transition probabilities between states E and NE are given in form of *logistic* functions. However, rather than using empirical data to estimate parameters of these functions, the authors parameterized the behavior diffusion model with reference to models in prior literature for other European cities. To determine adoption of water-saving technology, the opinion diffusion model is coupled with the technological diffusion model, which is implemented by a simple agent-based adaptation of the Bass model following Borshchev and Filippov (2004). The model was validated qualitatively by domain experts, quantitatively calibrated based on the first quarter of 2006, and validated by comparing the model with actual adoption in the following two quarters. The authors demonstrate that simulation results successfully replicate the consequence of a water-saving campaign on domestic water consumption.

Dugundji and Gulyás (2008) propose a computational model that combines econometric estimation with agent-based modeling to study the adoption of transportation options for households in a city in Netherlands. The presented discrete choice modeling framework aims to address interactions within different social and spatial network structures. Specifically, agent decision is captured using a *nested logit* model, which enables one to capture observed and unobserved behavior heterogeneity. Feedback effects among agents are introduced by adding a linear term (a so-called *field variable*) that captures proportions of an agent's neighbors making each decision to each agent's utility function. Because survey data on interactions between identifiable individuals was unavailable, this term only captured aggregate interactions among socioeconomic peers. The authors investigated simulated transition dynamics for the full model with two reference models: the first a nested logit model with a global field variable only and a fully connected network, and the second a multinomial logit model which is a special case to the full model. They found that simulated dynamics differ dramatically between the models. Given this lack of modeling robustness, no further validation was undertaken.

Tran (2012) develops an agent-based model to investigate energy innovation diffusion. Agent behavior in this model is determined by the relative importance of technology attributes to the agents, and social influence. Social influence, in turn, takes two forms: indirect influence coming from the general population, and direct influence of social network neighbors. The author drew on ABM studies in the marketing literature, and formulated the adoption model as $Prob(t) = 1 - (1 - P_{ij})(1 - Q_{ij})^{K_{ij}}$, where $P_{ij}$ captures individual choice using a discrete choice model of consumer decision-making, in which an agent's utility is defined as an inner product of coefficients and attributes. Coefficients are a random vector, with distribution different for different agents, capturing preference heterogeneity. $Q_{ij}$ and $K_{ij}$ is the indirect and direct network influence, respectively, captured as a function of the number of adopters at

decision time. While the model was evaluated using simulation experiments, and the nature of the model makes it well suited for empirically grounded parameter calibration, it was not in actuality quantitatively calibrated or validated using empirical data.

### 2.5.3 Machine learning models

*Machine learning* (ML) is a sub-area of computer science that aims to develop algorithms that uncover relationships in data. Within a supervised learning paradigm which is of greatest relevance here, the goal is further to develop models that accurately predict the value of an outcome variable for unseen instances. To do so, a computer program is expected to recognize patterns from a large set of observations, referred to as a *training* process that is grounded in statistical principles and governed by intelligent algorithms, and make predictions on new, unseen, instances. This category of methods has recently drawn much attention in academia and industry due to tremendous advances in predictive efficacy on important problems, such as image processing and autonomous driving. Combining machine learning with agent-based modeling seems promising in the study of innovation diffusion since the two can complement each other. The former is specialized in building a high-fidelity predictive models, while the latter captures dynamics and complex interdependencies. Of particular relevance to combining ML and ABM is the application of machine learning to model and predict human behavior. Interestingly, relatively few attempts have been made to date to incorporate ML-based models of human behavior within ABM simulations.

Sun and Müller (2013) develop an agent-based model that features Bayesian belief networks (BBNs) and opinion dynamics models (ODMs) to model land-use dynamics as they relate to payments for ecosystem services (PES). The decision model of each household is represented using a BBN, which were calibrated using survey data and based on discussions with relevant stakeholders, and incorporate factors such as income and land quality. Social interactions in decision-making are captured by ODM. The modeling framework was applied to evaluate China's Sloping Land Conversion Program (SLCP), considered among the largest PES programs. SLCP was designed to incentivize reforestation of land through monetary compensation. In their model, farmers make land-use decisions whether or not to participate in the SLCP program based on *internal* beliefs and *external* influences. External influences adjust internal beliefs cumulatively using a modified Deffuant model (Deffuant et al. 2002a) within a community-based small-world social network. Initial model structures were obtained using a structural learning algorithm, with results augmented using qualitative expert knowledge, resulting in a pseudo tree-augmented naive Bayesian (TAN) network. The final BBN model was validated by using a sensitivity analysis, and measuring prediction accuracy and area under the curve (AUC) of the receiver operating characteristics (ROC) curve on a holdout test data set at both household and plot level. A crucial limitation of this work is that only the BBN model was carefully validated; the authors did not validate the full simulation model at either the micro or macro levels.

Zhang et al. (2016) propose a data-driven agent-based modeling (DDABM) framework for modeling residential rooftop solar photovoltaic (PV) adoption in San Diego county. In this framework, the first step is to use machine learning to calibrate individual agent behavior based on data comprised of individual household characteristics and PV purchase decisions. These individual behavior models were validated using cross-validation methods to ensure predictive efficacy on data not used for model calibration, and were then used to construct an agent-based simulation with the learned model embedded in artificial

agents. In order to ensure validation on independent data, the entire time series data of individual adoptions was initially split along a time dimension. Training and cross-validation for developing the individual-level models were performed only on the first (early) portion of the dataset, and the aggregate model was validated by comparing its performance with actual adoptions on the second, independent time series, into the future relative to the calibration data set. The authors thereby rigorously demonstrate that the resulting agent-based model is effective in forecasting solar adoption both at the micro and macro levels. To our best knowledge, this work proposed the first generic principled framework that combines ML and ABM in study of innovation diffusion. Unlike most ABM studies we have reviewed, DDABM has the following features: (1) it does not make any assumptions on the structural features of social network, relying entirely on a data-driven process to integrate most predictive spatial and social influence features into the individual adoption model; (2) it does not rely on matching simulated dynamics with the empirical observations to calibrate the model, but instead parameterizes the model through a far more efficient statistical learning method at the level of individual agent behavior; and (3) validation is performed on *independent* data to evaluate the predictive effectiveness of the model. Moreover, validation is not only done at the macro-level by comparison with actual adoption traces, but also at the micro-level by means of the simulated *likelihood ratio* relative to a baseline model. To further justify the usefulness of ML-base approach, Zhang et al. (2016) actually implement and compare their model with another agent-based model of rooftop solar adoption developed by Palmer et al. (2015), with parameters calibrated on the same dataset following the general aggregate-level calibration approach used by them. The result is very revealing, as it strongly suggests that aggregate-level calibration is prone to overfit the model to data, an issue largely avoided by calibrating individual agent behavior.

## 2.6 Social influence models

Our last methodological category covers several models looking specifically at social influence. These models are quite simple, abstract, but prevalent in the theoretical study of innovation diffusion. Our purpose of discussing these is that there have been several recent efforts to calibrate these models using empirical data.

After analyzing an adoption dataset of Skype, Karsai et al. (2014) develop an agent-based model to predict diffusion of new online technologies. Specifically, agents in their model are characterized by three states: *susceptible* (S), *adopter* (A), and *removed* (R). Susceptible refers to people who may adopt the product later. Adopter agents have already adopted. Finally, removed are those who will not consider adopting the product in the future again. The transition from S to A is regulated by *spontaneous adoption* and *peer-pressure*, from A to S by *temporary termination*, and from A to R by *permanent termination*, each of which is parametrized by a constant probability which is identical for all users. While some parameters, such as average degree and temporary termination probability, are estimated directly from observations, the remaining parameters are determined by simultaneously fitting the empirical rates using a bounded nonlinear least-squares method. The model is fit over a 5-year training period, and validation uses predictions over the last six months of the observation period. However, validation is somewhat informal, since the predictability of the model is evaluated on a part of the training data and there is no validation of micro-behavior. In a later work using the same Skype data, Karsai et al. (2016) develop a threshold-driven social contagion model with only two states: susceptible and adopted. In addition, the model assumes that some fraction of nodes never adopt. The

authors calibrated the value of this fraction by matching the size of the largest component of adopters given by the simulations with real data. In addition, the model assumes that susceptible nodes adopt with a constant probability, which is informed by empirical analysis. In their simulations, nodes have heterogeneous degrees and thresholds, which follow empirical distributions. However, validation was not performed using independent data.

Herrmann et al. (2013) present two agent-based models of diffusion dynamics in online social networks. The first ABM is motivated by the Bass model, but time is discretized and each agent has two states: *unaware* and *aware*. At each time step, an unaware agent changes state to aware as a function of two triggers: innovation arising from exogenous sources, such as advertising, and imitation, which comes from observing decisions by neighbors. The second model termed the *independent cascade* model, originating from Goldenberg et al. (2001), has a similar structure to the agent-based Bass model, except that the imitation effect is formulated as a single probability with which each aware neighbor can independently change the state of an agent to *aware*. The author applied the two models in parallel to four diffusion data sets from Twitter, and calibrated parameters using actual aggregate adoption paths. Notably, validation is only performed at macro-level as an in-sample exercise, and shows that the two models behave similarly.

Using historical diffusion data of Facebook apps, Trusov et al. (2013) introduce an approach that applies Bayesian inference to determine a mixture of multiple network structures. Notice that most ABMs we reviewed so far either assume a single underlying social network (with parameters determined in model calibration) or generate artificial networks based on empirical findings or social science theories. They first choose a collection of feasible networks that represent the unobserved consumer networks. Then, a simple SIR model [similar to the Bass ABM in Herrmann et al. (2013)] is used to simulate the diffusion of products. The simulated time series are further transformed to multivariate stochastic functions, which provide priors to the Bayesian inference model to obtain the posterior weights on the set of feasible consumer networks. Like Herrmann et al. (2013), the adoption model is calibrated from the aggregate output, rather than from observations of individual decisions.

Chica and Rand (2017) propose an agent-based framework to build decision support system (DSS) for word-of-mouth programs. They developed a DSS to forecast the purchase of a freemium app and evaluate marketing policies, such as targeting and reward. The model captures seasonality of user activities by two probabilities for weekday and weekend respectively. The initial social network is generated by matching the degree distribution of the real network. Then, for each node, two weights are assigned to in- and out-edges, respectively, turning the network into a weighted graph that represents the heterogeneous social influence among social neighbors. Specifically, two models are used to model the information diffusion. One is the Bass-ABM (Rand and Rust 2011); the other is a contagion model (a threshold model but adding external influence). The parameters of the model were calibrated by a genetic algorithm (Stonedahl and Rand 2014), in which the fitness is defined based on the difference of simulated adoption from the historical adoption trajectory. Notably, the model was validated by a hold-out dataset, which is independent of the training data. For example, the entire 3 month period spanned by the data was divided into two: first 60 days for training, the last 30 days for validation.

The independent cascade model used by Herrmann et al. (2013) and the threshold model used by Chica and Rand (2017) are significant insofar as these connect to a substantial literature that has recently emerged within the Computer Science community on *information diffusion*, whereby information (broadly defined) spreads over a social network. We make this connection more precisely in Sect. 4 below.

## 3 Categorization of innovation diffusion models by application

Thus far, we followed a categorization of agent-based models of innovation diffusion focused on methods by which agent behavior is modeled. First, we observe that methods range from sophisticated mathematical optimization models (Sect. 2.1), to economics-based models (Sect. 2.2), to even simpler models based on heuristics for representing agent behavior (Sect. 2.4). While economic factors are dominant concerns in some applications, others emphasize the cognitive aspects of human decision-making (Sect. 2.3) and are frequently used to model influence over online social networks (Sect. 2.6). Second, we note that the method chosen to capture agent behavior also impacts the techniques used to calibrate model parameters from data. For example, cognitive models are often constructed based on detailed behavior data collected from field experiments and surveys, whereas models of agent behavior based on statistical principles rely on established statistical inference techniques for model calibration based on individual behavior data that is either observational or experimental. Other modeling approaches within our six broad categories often do not use data to calibrate individual agent behavior, opting instead to tune model parameters in order to match aggregate adoption data.

We now offer an alternative perspective to examine the literature on empirical ABMs of innovation diffusion by considering applications—that is, what particular innovation is being modeled. A breakup of existing work using this dimension is given in Table 2. As shown in the first column, we group applications by broad categories: agricultural innovations and farming, sustainable energy and conservation technologies, consumer technologies and innovations, information technologies and social goods. Interestingly, the first two categories account for more than half of the publications in literature. This likely reflects the history of ABM as an interdisciplinary modeling framework for computational modeling of issues that are of great interest in social science. A closely related factor could be the relatively high availability of data in these applications generated by social scientists (e.g., through the use of surveys). Another interesting observation that arises is methodological convergence for given applications: relatively few applications have been modeled within different methodological frameworks as categorized above. Future research may explore the use of different methods for same application. Furthermore, comparison of different modeling methods is rare within a single work [except in Dugundji and Gulyás (2008) and Zhang et al. (2016)], although such a methodological cross-validation is of importance as emphasized by some authors (Carley 1996; Rand and Rust 2011).

## 4 Information diffusion models

Online social networks have emerged as an crucial medium of communication. It does not only allow users to produce, exchange, and consume information at an unprecedented scale and speed, but also speeds the diffusion of novel and diverse ideas (Guille et al. 2013; Bakshy et al. 2012). The emergence of online social networks and advances in data science and machine learning have nourished a new field: *information diffusion*. The fundamental problem in information diffusion is to model and predict how information is propagated through interpersonal connections over social networks using large-scale diffusion data. In fact, several authors have reviewed the topic of information diffusion over online social networks (Bonchi 2011; Guille et al. 2013; Shakarian et al. 2015). Our aim is not to provide a comprehensive review of this same topic. Instead, we are interested in building connections between the agent-based modeling approach to innovation diffusion, and the modeling

**Table 2** Categorization of surveyed work by applications

| Category | Application | Method | Citation |
|---|---|---|---|
| Agricultural innovations and farming | Agricultural innovations | Mathematical programming | Berger (2001), Berger et al. (2007), Schreinemachers et al. (2007, 2009, 2010) and Alexander et al. (2013) |
| | | Economics (utility) | Holtz and Pahl-Wostl (2012) |
| | Organic farming | Cognitive model (deffuant) | Deffuant et al. (2002b) |
| | | Cognitive model (TPB, deffuant) | Kaufmann et al. (2009) |
| | Biogas plant | Economics (profit) | Sorda et al. (2013) |
| | Payments for ecosystem services | Machine learning | Sun and Müller (2013) |
| Sustainable energy and conservation technologies | Water-saving technology | Cognitive model (TPB) | Schwarz and Ernst (2009) |
| | | Discrete choice model | Galán et al. (2009) |
| | Heating system | Cognitive model (TPB) | Sopha et al. (2013) |
| | | Conjoint analysis | Lee et al. (2014) |
| | | Economics (cost) | Faber et al. (2010) |
| | Solar photovoltaic | Heuristic | Zhao et al. (2011) |
| | | Economics (utility) | Palmer et al. (2015) |
| | | Cognitive model (TPB, deffuant) | Rai and Robinson (2015) |
| | | Machine learning | Zhang et al. (2016) |
| | Fuel cell vehicles | Cognitive model (consumat) | Schwoon (2006) |
| | Energy innovation | Discrete choice model | Tran (2012) |
| | Electric vehicles | Cognitive model (TEC) | Wolf et al. (2012) |
| | | Economics (utility) | Plötz et al. (2014) |
| | | Economics (utility) | McCoy and Lyons (2014) |
| | Alternative fuel vehicles | Conjoint analysis | Zhang et al. (2011) |
| | Alternative fuels | Heuristic | Van Vliet et al. (2010) |
| | | Economics (utility) | Günther et al. (2011) |
| | | Canalysis | Stummer et al. (2015) |
| | Green electricity | Cognitive model (LARA) | Krebs and Ernst (2015) and Ernst and Briegel (2017) |
| | Air-quality feedback device | Cognitive model (TPB) | Jensen et al. (2016) |
| Consumer technologies and innovations | Wine bottle closures | Conjoint analysis | Garcia et al. (2007) |
| | Mobile phones | Conjoint analysis | Vag (2007) |
| | Transportation mode | Discrete choice model | Dugundji and Gulyás (2008) |
| | New cars | Fuzzy TOPSIS (heuristic) model | Kim et al. (2011) |

**Table 2** continued

| Category | Application | Method | Citation |
|---|---|---|---|
| | Movie | Economics (utility) | Broekhuizen et al. (2011) |
| Information technologies | Skype | Social contagion model | Karsai et al. (2014, 2016) |
| | Twitter | Independent cascade model | Herrmann et al. (2013) |
| | Facebook app | Social contagion model | Trusov et al. (2013) |
| | Freemium app | Social contagion model | Chica and Rand (2017) |
| Social goods | Neighborhood support | Cognitive model (LARA) | Krebs et al. (2013) |

methods in the field of information diffusion. Indeed, researchers in the ABM community have paid little attention to the existing methods for modeling information diffusion, and especially in the played by data science in this field, which has significant implications for ABM model calibration, as we discuss below.

### 4.1 Two basic models of information diffusion

Compared to agent adoption models in Sect. 2, the decision process in the information diffusion literature is typically very simple, following predominantly the social influence models. The two most common models in information diffusion are Independent Cascades (IC) (Goldenberg et al. 2001) and Linear Threshold (LT) models (Granovetter 1978). These models are defined on directed graphs where activation is assumed to be monotonic: once a node is active (e.g., adopted, received information), it cannot become inactive. The diffusion process in both models starts with a few active nodes and progresses iteratively in a *discrete* and *synchronous* manner until no new nodes can be infected. Specifically, in each iteration, a new active node in the IC model is given a single chance to activate its inactive neighbors independently with an exogenously specified probability (usually represented by the weight of the corresponding edge). In the LT model, in contrast, an inactive node will become active only if the sum of weights of its activated neighbors exceeds a predefined node-specific threshold, which is typically randomly assigned between 0 and 1 for each network node. Note that in both models a newly activated node becomes active immediately in the next iteration. From an agent-based perspective, both IC and LT are generative models which define two diffusion mechanisms.

### 4.2 Learning information diffusion models

Several efforts use empirical data to calibrate the parameters of the LT and IC models. Saito et al. (2011) propose an asynchronous IC (AsIC) model, which not only captures temporal dynamics, but also node attributes. They show how the model parameters can be estimated from observed diffusion data using maximum likelihood estimation (MLE). The AsIC model closely follows the IC model, but additionally introduces a time delay before a newly activated node becomes active. The time delay is assumed to be exponentially distributed with a parameter that is defined as an exponential function of a feature vector (a composition of attributes associated with both nodes and edges). The transmission probability is then defined as a logit function of the feature vector. The data is given in the format of "diffusion traces", and each trace is a sequence of tuples which specify activation time for a subset of nodes.

To learn the model using this data, the authors define the log-likelihood of the data given the model. The authors then demonstrate how to solve the resulting optimization problem using expectation-maximization (EM). While the proposed model is promising to be used for prediction, the learning method was only tested using synthetic data.

Guille and Hacid (2012) show how to parameterize the AsIC model using machine learning methods based on Twitter data. In their model, the diffusion probability for information at any given time between two users is a function of attributes from three dimensions: *social*, *semantic*, and *time*, which group features with respect to social network, content and temporal property respectively. Four different classifiers were trained and compared in terms of cross-validation error: C4.5 decision tree, linear perceptron, multilayer perceptron, and Bayesian logistic regression. The last model mentioned above was finally used for prediction. Notably, time-delay parameter was determined separately in this work by comparing simulation results with actual diffusion dynamics, which is the same calibration method commonly used in ABM of innovation diffusion. Unlike (Saito et al. 2011), where all model parameters are inferred by MLE, here only a subset of model parameters are estimated through established machine learning techniques, but the rest are calibrated by simulations. Their evaluation shows that the model accurately predicts diffusion dynamics, but fails to accurately predict the volume of tweets. In our ABM jargon, the model performs well at macro-level, but poorly at micro-level validation (Carley 1996). Another limitation of this work is that validation is only performed as an in-sample exercise, rather than using out-of-sample data.

Galuba et al. (2010) propose two diffusion models with temporal features that are used to predict user re-tweeting behaviors on Twitter. Both models define the probability for a user to re-tweet a given URL to be a product of two terms: one is *time-independent*, the other is *time-dependent*. Both have the same time-dependent part which follows a log-normal distribution, but differ in the actual definitions of the time-independent part. In their first model termed At-Least-One (ALO), the time-independent component is defined as the likelihood of at least one of the causes: either one is affected by the agent it follows, or by the user tweets a URL spontaenously. The second, Linear Threshold (LT), model, posits that a user re-tweets a URL only if the cumulative influence from all the followees is greater than a threshold. The time-independent component in this model is given by a sigmoid function. In order to calibrate and validate the model, the data set was split along the time dimension into two parts. The model was calibrated by choosing parameters that optimize the estimated F-score using the gradient ascent method on the first (earlier) data set, and used to predict URL mentions in the second (later) data set. Their results show that the LT model achieves the highest F-score among all models and correctly predicts approximately half of URL mentions with lower than 15% false positives.

While all research reviewed so far assumes known network structure, a number of efforts deal with hidden network structures which must also be learned from data. The so-called *network inference problem* is to infer the underlying network given a complete activation sequence (Guille et al. 2013). Gomez Rodriguez et al. (2010) introduce a variant of the independent cascade model (Kempe et al. 2003) adding time delay. Their problem is to find a directed graph with at most $k$ edges that maximizes the likelihood of a set of cascades for a given transmission probability and parameters of the incubation distribution, which is solved approximately using a greedy algorithm. Myers and Leskovec (2010) propose a cascade model which is similar to Gomez Rodriguez et al. (2010) but allows distinct transmission probabilities for different network edges. The goal is to infer the adjacency matrix (referring to the pairwise transmission probabilities) that maximizes the likelihood given a set of cascades, which is accomplished by solving a convex optimization problem derived from the problem formulation. Gomez Rodriguez et al. (2011) develop a continuous-time diffusion model that

unifies the two-step diffusion process involving both a transmission probability and time delay from Gomez Rodriguez et al. (2010) and Myers and Leskovec (2010). The pivotal value is the *conditional* probability for a node $i$ to be infected at time $t_i$ given that a neighboring node $j$ was infected at time $t_j$, which is formulated as a function of the time interval $(t_i - t_j)$ and parametrized by a pairwise transmission rate $\alpha_{ji}$. Survival analysis (Lawless 2011) is used to derive the maximum likelihood function given a set of cascades, and they aim to find a configuration of all transmission rates that maximizes the likelihood. While most network inference algorithms assume static diffusion networks, Gomez Rodriguez et al. (2013) address a network inference problem with a time-varying network. The resulting inference problem is solved using an online algorithm upon formulating the problem as a stochastic convex optimization.

### 4.3 Bridging information diffusion models and agent-based modeling of innovation diffusion

The methodological framework of the information diffusion inference problems discussed above is a natural fit for principled data-driven agent-based modeling. The information diffusion models characterized by transmission probabilities and time delay are essentially agent-based models. Given data of diffusion cascades, they can be constructed either using only the temporal event (adoption) sequence, or using more general node features, social network, content, and any other explanatory or predictive factors. In fact, ABM researchers have started to apply similar statistical methods to develop empirical models (see Sect. 2.5). Notably, as shown by Zhang et al. (2016), parametric probabilistic models of agent behavior can be estimated from observation data using maximum likelihood estimation methods. In addition, the approaches for network inference appear particularly promising in estimating not only behavior for a known, fixed social influence network, but for estimating the influence network itself, as well as the potentially heterogeneous influence characteristics.

A crucial challenge in translating techniques from information diffusion domains to innovation diffusion is that in the latter only observes a single, partial adoption sequence, rather than a collection of complete adoption sequences over a specified time interval. As a consequence, the fully heterogeneous agent models cannot be inferred, although the likelihood maximization can still be effectively formulated by limiting the extent of agent heterogeneity [with the limit of homogeneous agents used by Zhang et al. (2016)]. In addition, the assumptions generally made in information diffusion models can also pose serious challenges to the transferability of the approach to agent-based modeling. Recall that the information cascade models often assume that an adopter has a *single* chance to affect its inactive neighbors and a non-adopter is affected by its neighboring adopters *independently*. These assumptions simplify the construction of the likelihood function, but further justification is needed for them, especially when building empirical models that are expected to faithfully represent realistic social systems and diffusion processes. Note that rules that govern the interactions in agent-based models are quite flexible and can be very sophisticated, which is also one of the major advantages of agent-based computing over analytical models. Although one may be able to explicitly derive a parametric likelihood function given diffusion traces in more complex settings than existing information diffusion models do, this is sure to be technically challenging. Moreover, solving the resulting MLE can be computationally intractable. Therefore, to take advantage of MLE approach in information diffusion, ABM researchers must make appropriate assumptions on agent interactions so that they can derive tractable likelihood functions without significantly weakening the model's explanatory and predictive power.

## 5 Discussion

### 5.1 Validation in agent-based modeling

As agent-based modeling is increasingly called for in service of decision support and prediction, it is natural to expect them to be empirically grounded. An overarching consideration in empirically grounded agent-based modeling is how data can be used in order to develop reliable models, where reliability is commonly identified with their ability to accurately represent or predict the environment being modeled. This property of *reliability* is commonly confirmed through model *validation*. In social science, a number of authors have contributed to the topic of validation, from approaches for general computational models (Carley 1996), to those focused on agent-based simulations (Xiang et al. 2005; Fagiolo et al. 2006; Garcia et al. 2007; Ormerod and Rosewell 2009; Rand and Rust 2011), to specific types of agent-based models (Brown et al. 2005). Outside of social science, validation of simulation systems has an even longer history of investigation (Knepell and Arangno 1993; Banks 1998; Kleijen 1999; Sanchez 2001). We now briefly review these approaches.

As previously mentioned, Carley (1996) suggests four levels of validation: *grounding*, *calibration*, *verification*, and *harmonizing*. *Grounding* establishes reasonableness of a computational model, including face validity, parameter validity, and process validity; *calibration* establishes model's feasibility by tuning a model to fit empirical data; *verification* demonstrates how well a model's predictions match data; and *harmonization* examines the theoretical adequacy of a verified computational model.

More recently, drawing on formal model verification and validation techniques from industrial and system engineering for discrete-event system simulations, Xiang et al. (2005) suggest the software implementation of agent-based model has to be verified with respect to its conceptual model, and highlight a selection of validation techniques from Banks (1998), such as face validation, internal validation, historical data validation, parameter variability, predictive validation, and Turing tests. Moreover, they suggest the use of other complementary techniques, such as model-to-model comparison (Axtell et al. 1996) and statistical tests (Kleijen 1999; Sanchez 2001).

For agent-based models in economics, Fagiolo et al. (2006) proposed three different types of *calibration* methods: the *indirect calibration* approach, the *Werker–Brenner empirical calibration* approach, and the *history-friendly* approach. For example, Garcia et al. (2007) adopt the last approach to an innovation diffusion study in New Zealand winery industry, using conjoint analysis to instantiate, calibrate, and verify the agent-based model qualitatively using stylized facts.

For agent-based models in marketing, Rand and Rust (2011) suggest verification and validation as two key processes as guidelines for rigorous agent-based modeling. The use of term "verification" follows common understanding in system engineering (Xiang et al. 2005). In particular, the authors identify four steps for validation: micro-face validation, macro-face validation, empirical input validation, and empirical output validation using stylized facts, real-world data, and cross-validation. Note that the proposed validation steps echo the framework by Carley (1996): the first two steps correspond to grounding, the third to calibration, and the fourth roughly combines verification and harmonization. However, the cross-validation method mentioned in Rand and Rust (2011) appears to suggest validation across models, whereas Carley (1996) suggests validation across multiple data sets. The latter is consistent with the use of cross-validation in statistical inference and machine learning (Friedman et al. 2001; Bishop 2006).

Focusing specifically on empirically grounded ABMs, we suggest two pivotal steps in ensuring model reliability in a statistical sense: *calibration* and *validation*. By calibration, we mean the process of *quantitatively* fitting a set of model parameters to data, whereas validation means a quantitative assessment of the predictive efficacy of the model *using independent data*, that is, using data which was not utilized during the calibration step. Moreover, insofar as a model of innovation diffusion is concerned with predicting future diffusion of an innovation, we propose to further split the dataset along a temporal dimension, so that earlier data is used exclusively for model calibration, while later data exclusively for validation. Starting with this methodological grounding, we now proceed to identify common issues that arise in prior research on empirically grounded agent-based models of innovation diffusion.

## 5.2 Issues in model calibration and validation

Agent-based modeling research has often been criticized for lack of accepted methodological standard, hindering its acceptance in top journals by mainstream social scientists. One notable protocol due to Richiardi et al. (2006) highlight four potential methodological pitfalls: *link with the literature*, *structure of the models*, *analysis*, and *replicability*.

A careful examination of the empirical ABM work on innovation diffusion through this protocol suggests that most of these issues have been addressed or significantly mitigated. For example, nearly all of the reviewed papers present theoretical background, related work, sufficient description of model structure, sensitivity analysis of parameter variability, a formal representation (e.g., UML,[3] OOD[4]), and public access to source code. In spite of these improvements, however, there are residual concerns about systematic quantitative calibration and validation using empirical data.

We observe that different agent adoption models are calibrated differently. In the case of cognitive agent models (Sect. 2.3), such as the Theory of Planned Behavior and theory of emotional coherence, the individual model parameters are often estimated using survey data. Similarly, statistics-based models (Sect. 2.5) can be parametrized using either experimental or observational individual-level data. On the other hand, for conceptual models, such as heuristic (Sect. 2.4) and economics-based models (Sect. 2.2), calibration is commonly done by iteratively adjusting parameters to match simulated diffusion trajectory to aggregate-level empirical data. Formally, we call the first kind of calibration "*micro-calibration*", as it uses individual data during calibration, whereas the second type "*macro-calibration*", as it uses aggregate-level data. Moreover, in many studies simulation parameters are determined using both micro- and macro-calibration. For example, since network structure is often not fully observed, and rules that govern agent interactions are assumed, parameters of these are commonly macro-calibrated. Our first concern is about macro-calibration.

### 5.2.1 Issue I: potential pitfalls in macro-calibration

When a model has many parameters, over-fitting the model to data becomes a major concern (Friedman et al. 2001; Bishop 2006). As Carley (1996) suggests, "any model with sufficient parameters can always be adjusted so that some combination of parameters generates the observed data, therefore, large multi-parameter models often run the risk of having so many parameters that there is no guarantee that the model is doing anything more than curve fitting."

---

[3] The short for the Unified Modeling Language, developed by the Object Management Group: http://www.omg.org.

[4] A standard to describe agent-based models originally proposed by Grimm et al. (2006) for ecological modeling.

Interestingly, the issue of over-fitting may even be a concern in macro-calibration when only a few parameters need to be calibrated. The reason is that agent-based models are highly non-linear, and even small changes in several parameters can give rise to substantially different model dynamics. This issue is further exacerbated by the fact that macro-calibration makes use of aggregate-level data, which is often insufficient in scale for reliable calibration of any but the simplest models, as many parameter variations can give rise to similar aggregate dynamics.

Addressing the issue requires greater care and rigor in applying macro-calibration. One possibility is that instead of choosing only a single parameter configuration, to select a parameter *zone* using a classifier such as decision trees (Kaufmann et al. 2009) or other machine learning algorithms. Subsequently, the variability of parameters within this zone can be further investigated using sensitive analysis. Another potential remedy is that instead of using only a single target statistic (e.g., average adoption rates) to use multiple indicators. A relevant strategy to build agent-based models in the field of ecology is termed "pattern-oriented modeling", which utilizes multiple patterns at different scales and hierarchical levels observed from real systems to determine the model structure and parameters (Grimm et al. 2005).

In addition, there are more advanced and robust techniques that can improve the rigor of macro-calibration. The modeling framework in Zhang et al. (2016) and statistical inference methods introduced in Sect. 4 propose methods which integrate micro and macro calibration into a single maximum likelihood estimation framework. Through well-established methods in machine learning, such as cross-validation, one can expect to parameterize a highly-predictive agent-based model and minimize the risk of over-fitting. Indeed, a fundamental feature of any approach should be to let validation ascertain the effectiveness of macro-calibration in generalizing beyond the calibration dataset. This brings us to the second common issue revealed by our review: lack of validation on independent data.

### 5.2.2 Issue II: rigorous quantitative validation on independent data is uncommon

A common issue in the research we reviewed is that validation is often informal, incomplete, and even missing. The common reason for incomplete data-driven validation is that relevant data is simply unavailable. However, so long as data is available for calibrating the model, one can in principle use this data for both calibration and validation steps, for example, following cross-validation methods commonly utilized in machine learning. Several efforts seek to standardize the validation process for agent-based models, and computational models in general. However, few papers discussed explicitly follow any formalized validation approaches in this literature, although important exceptions exist (Garcia et al. 2007; Zhang et al. 2011; Stummer et al. 2015).

### 5.2.3 Issue III: few conduct validation at both micro-level and macro-level

There has been some debate about whether validation should be performed at both micro- and macro-level (Carley 1996). While arguments against the dual-verification often emphasize greater importance of model accuracy at the aggregate level, we argue that robust predictions at the aggregate level can only emerge when individual behavior is accurately modeled as well, particularly when policies that the ABM evaluates can be implemented as modifying individual decisions.

Statistics-based models, such as machine learning, have well-established validation techniques which can be leveraged to validate individual-level models. One widely-used

technique in machine learning and data mining is *cross-validation*. A common use of cross-validation is by partitioning the data into $k$ parts, with training performed on $k - 1$ of these and testing (evaluation) on the $k$th. The results are then averaged over $k$ independent runs using each of the parts as test data. Observe that such a cross-validation approach can be used for models of individual behavior that are not themselves statistically-driven, such as models based on the theory of planned behaviors. Unfortunately, few of the surveyed papers, with the exception of statistics-based models, use cross-validation.

### 5.2.4 Issue IV: few conduct validation of forecasting effectiveness on independent "future" data

One limitation of cross-validation techniques as traditionally used is that they provide an offline assessment of model effectiveness. To assess the predictive power of dynamical systems, the entire model has to be validated in terms of its ability to predict "future" data relative to what was used in calibration. We call this notion "*forward validation*". In particular, forward validation must assess simulated behaviors against empirical observations at both individual and aggregate levels *with an independent set of empirical data*. This can be attained, for example, by splitting a time-stamped data set so that calibration is performed on data prior to a split date, and forward validation is done on data after the split date (Galán et al. 2009; Zhang et al. 2016; Rai and Robinson 2015; Chica and Rand 2017). In this review, we do observe several approaches that are validated on independent data, but these either are not looking forward in time relative to the calibration data, or only focus on macro-level validation. A common argument for the use of in-sample data for the forward validation is that new data is not available while the modeling task is undertaken. Notice, however, that any data set that spans a sufficiently long period of time can be split along the time dimension as above to effect rigorous forward validation.

### 5.3 Recommended techniques for model calibration and validation

We have identified several issues in calibration and validation which commonly arise in the prior development of empirical agent-based models for innovation diffusion, and briefly discussed possible techniques that can help address these issues. We now summarize our recommendations:

*Multi-indicator calibration* When macro-calibration is needed, the use of multiple indicators can help address *over-fitting*, whereby a model which appears to effectively *match* data in calibration performs poorly in prediction on unseen data. We suggest that such indicators are developed at different scale and hierarchical levels, so that models which cannot effectively generalize to unseen data can be efficiently eliminated.

*Maximum likelihood estimation* When individual-level data are available, we recommend constructing probabilistic adoption models for agents, and estimating parameters of these models by maximizing a global likelihood function [see, for example, the modeling framework by Zhang et al. (2016), and research discussed in Sect. 4.2]. Doing so offers a principled means of calibrating agent behavior models from empirical data.

*Cross validation* This approach is widely used for model selection in the machine learning literature. Here, we recommend it for both micro-calibration and micro-validation of ABMs. Note that it does not only apply to statistics-based models, but can be used for any agent modeling paradigm where model parameters are calibrated using empirical data. The use of cross-validation in calibration can dramatically reduce the risk of over-

fitting. Moreover, as it inherently uses independent data, such validation leads to more rigorous ABM methodology.

*Forward validation* This method involves splitting data into two consecutive time periods. The modeler calibrates an agent-based model using data from the first period, and assesses the predictive efficacy of the model in the second period. More rigorously, validation of the model should be evaluated at both individual and aggregate levels.

# 6 Conclusions

We provided a systematic, comprehensive, and critical review of existing work on empirically grounded agent-based models for innovation diffusion. We offered a unique methodological survey of literature by categorizing agent adoption models along two dimensions: methodology and application. We identified six methodological categories: *mathematical optimization based models*, *economics-based models*, *cognitive agent models*, *heuristic models*, *statistics-based models* and *social influence models*. They differ not only in terms of assumptions and elaborations of human decision-making process, but also with respect to calibration and parameterization techniques. Our critical assessment of each work focused on using data for calibration and validation, and particularly performing validation with independent data. We briefly reviewed the most important work in the closely related literature on information diffusion, building connections between the innovation and information diffusion approaches. One particularly significant observation is that information diffusion methods rely heavily on machine learning and maximum likelihood estimation approaches, and the specific methodology used can be naturally ported to innovation diffusion ABMs. Drawing on prior work in validation of computational models, we discussed four main issues for existing empirically grounded ABM studies in innovation diffusion, and provided corresponding solutions.

On balance, recent developments of empirical approaches in agent-based modeling for innovation diffusion are encouraging. Although calibration and validation issues remain in many studies, a number of natural solutions from data analytics offer promising directions in this regard. The ultimate goal of empirically grounded ABMs is to provide decision support for policy makers and stakeholders across a broad variety of innovations, helping improve targeted marketing strategies, and reduce costs of successful translation of high-impact innovative technologies to the marketplace.

# References

Ajzen I (1991) The theory of planned behavior. Organ Behav Hum Decis Process 50(2):179–211

Alexander P, Moran D, Rounsevell MD, Smith P (2013) Modelling the perennial energy crop market: the role of spatial diffusion. J R Soc Interface 10(88):20130,656

Alexander P, Moran D, Smith P, Hastings A, Wang S, Sünnenberg G, Lovett A, Tallis MJ, Casella E, Taylor G et al (2014) Estimating uk perennial energy crop supply using farm-scale models with spatially disaggregated data. GCB Bioenergy 6(2):142–155

Axelrod RM (1997) The complexity of cooperation: agent-based models of competition and collaboration. Princeton University Press, Princeton

Axtell R, Axelrod R, Epstein JM, Cohen MD (1996) Aligning simulation models: a case study and results. Comput Math Organ Theory 1(2):123–141

Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. In: International world wide web conference, ACM, pp 519–528

Banks J (1998) Handbook of simulation: principles, methodology, advances, applications, and practice. Wiley, New York

Barabási AL, Albert R, Jeong H (1999) Mean-field theory for scale-free random networks. Phys A Stat Mech Appl 272(1):173–187

Bass FM (1969) A new product growth for model consumer durables. Manag Sci 15(5):215–227

Berger T (2001) Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. Agric Econ 25(2–3):245–260

Berger T, Birner R, Mccarthy N, DíAz J, Wittmer H (2007) Capturing the complexity of water uses and water users within a multi-agent framework. Water Resour Manag 21(1):129–148

Bishop CM (2006) Pattern recognition and machine learning. Springer, Berlin

Bonchi F (2011) Influence propagation in social networks: a data mining perspective. IEEE Intell Inform Bull 12(1):8–16

Borshchev A, Filippov A (2004) From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools. In: International conference of the system dynamics society, Citeseer, vol 22

Briegel R, Ernst A, Holzhauer S, Klemm D, Krebs F, Piñánez AM (2012) Social-ecological modelling with lara: a psychologically well-founded lightweight agent architecture. PhD thesis, International Environmental Modelling and Software Society (iEMSs)

Broekhuizen TL, Delre SA, Torres A (2011) Simulating the cinema market: how cross-cultural differences in social influence explain box office distributions. J Prod Innov Manag 28(2):204–217

Brown DG, Page S, Riolo R, Zellner M, Rand W (2005) Path dependence and the validation of agent-based spatial models of land use. Int J Geogr Inf Sci 19(2):153–174

Carley KM (1996) Validating computational models. http://www.casos.cs.cmu.edu/publications/papers.php

Chandrasekaran D, Tellis GJ (2007) A critical review of marketing research on diffusion of new products. Rev Market Res 3(1):39–80

Chica M, Rand W (2017) Building agent-based decision support systems for word-of-mouth programs: a freemium application. J Market Res. doi:10.1509/jmr.15.0443 (in press)

Dawid H (2006) Agent-based models of innovation and technological change. Handb Comput Econ 2:1235–1272

Deffuant G, Neau D, Amblard F, Weisbuch G (2000) Mixing beliefs among interacting agents. Adv Complex Syst 3(01n04):87–98

Deffuant G, Amblard F, Weisbuch G, Faure T (2002a) How can extremism prevail? A study based on the relative agreement interaction model. J Artif Soc Soc Simul 5(4). http://jasss.soc.surrey.ac.uk/5/4/1.html

Deffuant G, Huet S, Bousset JP, Henriot J, Amon G, Weisbuch G et al (2002b) Agent based simulation of organic farming conversion in Allier département. Complex Ecosyst Manag 158–189

Dugundji ER, Gulyás L (2008) Sociodynamic discrete choice on networks in space: impacts of agent heterogeneity on emergent outcomes. Environ Plan B Plan Des 35(6):1028–1054

Epstein JM (1999) Agent-based computational models and generative social science. Complexity 4(5):41–60

Ernst A, Briegel R (2017) A dynamic and spatially explicit psychological model of the diffusion of green electricity across Germany. J Environ Psychol 52:183–193

Faber A, Valente M, Janssen P (2010) Exploring domestic micro-cogeneration in the netherlands: an agent-based demand model for technology diffusion. Energy Policy 38(6):2763–2775

Fagiolo G, Windrum P, Moneta A (2006) Empirical validation of agent-based models: a critical survey. Technical report LEM working paper series

Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning. Springer, Berlin

Galán JM, López-Paredes A, Del Olmo R (2009) An agent-based model for domestic water management in valladolid metropolitan area. Water Resour Res 45(5). doi:10.1029/2007WR006536

Galuba W, Aberer K, Chakraborty D, Despotovic Z, Kellerer W (2010) Outtweeting the twitterers-predicting information cascades in microblogs. WOSN 10:3–11

Garcia R (2005) Uses of agent-based modeling in innovation/new product development research*. J Prod Innov Manag 22(5):380–398

Garcia R, Jager W (2011) From the special issue editors: agent-based modeling of innovation diffusion*. J Prod Innov Manag 28(2):148–151

Garcia R, Rummel P, Hauser J (2007) Validating agent-based marketing models through conjoint analysis. J Bus Res 60(8):848–857

Gilbert N, Troitzsch K (2005) Simulation for the social scientist. McGraw-Hill Education, New York

Goldenberg J, Libai B, Muller E (2001) Talk of the network: a complex systems look at the underlying process of word-of-mouth. Market Lett 12(3):211–223

Gomez Rodriguez M, Leskovec J, Krause A (2010) Inferring networks of diffusion and influence. In: ACM SIGKDD international conference on knowledge discovery and data mining, ACM, pp 1019–1028

Gomez Rodriguez M, Balduzzi D, Schölkopf B, Scheffer GT, et al (2011) Uncovering the temporal dynamics of diffusion networks. In: International conference on machine learning, international machine learning society, pp 561–568

Gomez Rodriguez M, Leskovec J, Schölkopf B (2013) Structure and dynamics of information pathways in online media. In: ACM international conference on web search and data mining, ACM, pp 23–32

Granovetter M (1978) Threshold models of collective behavior. Am J Sociol 83(6):1420–1443

Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke HH, Weiner J, Wiegand T, DeAngelis DL (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310(5750):987–991

Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, Giske J, Goss-Custard J, Grand T, Heinz SK, Huse G et al (2006) A standard protocol for describing individual-based and agent-based models. Ecol Model 198(1):115–126

Guille A, Hacid H (2012) A predictive model for the temporal dynamics of information diffusion in online social networks. In: International world wide web conference, ACM, pp 1145–1152

Guille A, Hacid H, Favre C, Zighed DA (2013) Information diffusion in online social networks: a survey. ACM SIGMOD Rec 42(2):17–28

Günther M, Stummer C, Wakolbinger L, Wildpaner M (2011) An agent-based simulation approach for the new product diffusion of a novel biomass fuel. J Oper Res Soc 62(1):12–20

Hauser J, Tellis GJ, Griffin A (2006) Research on innovation: a review and agenda for marketing science. Market Sci 25(6):687–717

Hegselmann R, Krause U et al (2002) Opinion dynamics and bounded confidence models, analysis, and simulation. J Artif Soc Soc Simul 5(3). http://jasss.soc.surrey.ac.uk/5/3/2.html

Herrmann J, Rand WM, Schein B, Vodopivec N (2013) An agent-based model of urgent diffusion in social media. Robert H Smith School research paper

Holland JH (1995) Hidden order: how adaptation builds complexity. Basic Books, New York

Holtz G, Pahl-Wostl C (2012) An agent-based model of groundwater over-exploitation in the Upper Guadiana, Spain. Reg Environ Change 12(1):95–121

Hopp WJ (2004) Ten most influential papers of management science's first 50 years. Manag Sci 50(12–supplement):1763–1763

Jager W (2000) Modelling consumer behaviour. Rijksuniversiteit Groningen, Groningen

Jager W, Janssen M, De Vries H, De Greef J, Vlek C (2000) Behaviour in commons dilemmas: homo economicus and homo psychologicus in an ecological-economic model. Ecol Econ 35(3):357–379

Janssen MA, Jager W (2002) Stimulating diffusion of green products. J Evolut Econ 12(3):283–306

Jensen T, Holtz G, Baedeker C, Chappin ÉJ (2016) Energy-efficiency impacts of an air-quality feedback device in residential buildings: an agent-based modeling assessment. Energy Build 116:151–163

Karsai M, Iñiguez G, Kaski K, Kertész J (2014) Complex contagion process in spreading of online innovation. J R Soc Interface 11(101):20140,694

Karsai M, Iñiguez G, Kikas R, Kaski K, Kertész J (2016) Local cascades induced global contagion: how heterogeneous thresholds, exogenous effects, and unconcerned behaviour govern online adoption spreading. Sci Rep 6. doi:10.1038/srep27178

Kaufmann P, Stagl S, Franks DW (2009) Simulating the diffusion of organic farming practices in two new eu member states. Ecol Econ 68(10):2580–2593

Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. In: ACM SIGKDD international conference on knowledge discovery and data mining, ACM, pp 137–146

Kiesling E, Günther M, Stummer C, Wakolbinger LM (2012) Agent-based simulation of innovation diffusion: a review. Cent Eur J Oper Res 20(2):183–230

Kim S, Lee K, Cho JK, Kim CO (2011) Agent-based diffusion model for an automobile market with fuzzy topsis-based product adoption process. Expert Syst Appl 38(6):7270–7276

Kleijen J (1999) Validation of models: statistical techniques and data availability. In: Proceedings of the 31st conference on winter simulation: simulation—a bridge to the future, vol 1, pp 647–654

Knepell PL, Arangno DC (1993) Simulation validation: a confidence assessment methodology, vol 15. Wiley, New York

Krebs F, Ernst A (2015) A spatially explicit agent-based model of the diffusion of green electricity: model setup and retrodictive validation. In: Social simulation conference, european social simulation association

Krebs F, Holzhauer S, Ernst A (2013) Modelling the role of neighbourhood support in regional climate change adaptation. Appl Spat Anal Policy 6(4):305–331

Lawless JF (2011) Statistical models and methods for lifetime data, vol 362. Wiley, New York

Lee T, Yao R, Coker P (2014) An analysis of uk policies for domestic energy reduction using an agent based tool. Energy Policy 66:267–279

Lewin R (1999) Complexity: life at the edge of chaos. University of Chicago Press, Chicago

Macal CM (2016) Everything you need to know about agent-based modelling and simulation. J Simul 10(2):144–156

Macal C, North M (2010) Tutorial on agent-based modelling and simulation. J Simul 4(3):151–162

Macy MW, Willer R (2002) From factors to actors: computational sociology and agent-based modeling. Ann Rev Sociol 28:143–166

Mahajan V, Muller E, Bass FM (1990) New product diffusion models in marketing: a review and directions for research. J Market 54(1):1–26

Mahajan V, Muller E, Wind Y (2000) New-product diffusion models, vol 11. Springer, Berlin

Matthews RB, Gilbert NG, Roach A, Polhill JG, Gotts NM (2007) Agent-based land-use models: a review of applications. Landsc Ecol 22(10):1447–1459

McCoy D, Lyons S (2014) Consumer preferences and the influence of networks in electric vehicle diffusion: an agent-based microsimulation in ireland. Energy Res Soc Sci 3:89–101

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27:415–444

Meade N, Islam T (2006) Modelling and forecasting the diffusion of innovation-a 25-year review. Int J Forecast 22(3):519–545

Michalek JJ, Papalambros PY, Skerlos SJ (2004) A study of fuel efficiency and emission policy impact on optimal vehicle design decisions. J Mech Des 126(6):1062–1070

Myers S, Leskovec J (2010) On the convexity of latent social network inference. In: Advances in neural information processing systems, pp 1741–1749

Negahban A, Yilmaz L (2014) Agent-based simulation applications in marketing research: an integrated review. J Simul 8(2):129–142

Noble J, Davy S, Franks DW (2004) Effects of the topology of social networks on information transmission. In: International conference on simulation of adaptive behavior, pp 395–404

Ormerod P, Rosewell B (2009) Validation and verification of agent-based models in the social sciences. In: Squazzoni F (ed) Epistemological aspects of computer simulation in the social sciences: second international workshop, EPOS 2006, Brescia, Italy, October 5–6, 2006, Revised Selected and Invited Papers, Springer, Berlin (pp. 130–140). doi:10.1007/978-3-642-01109-2_10

Palmer J, Sorda G, Madlener R (2015) Modeling the diffusion of residential photovoltaic systems in italy: an agent-based simulation. Technol Forecast Soc Change 99:106–131

Peres R, Muller E, Mahajan V (2010) Innovation diffusion and new product growth models: a critical review and research directions. Int J Res Market 27(2):91–106

Plötz P, Gnann T, Wietschel M (2014) Modelling market diffusion of electric vehicles with real world driving datapart I: model structure and validation. Ecol Econ 107:411–421

Rai V, Robinson SA (2015) Agent-based modeling of energy technology adoption: empirical integration of social, behavioral, economic, and environmental factors. Environ Model Softw 70:163–177

Rand W, Rust RT (2011) Agent-based modeling in marketing: guidelines for rigor. Int J Res Market 28(3):181–193

Richiardi M, Leombruni R, Saam NJ, Sonnessa M (2006) A common protocol for agent-based social simulation. J Artif Soc Soc Simul 9(1):15

Rogers EM (1995) Diffusion of innovations. Free Press, New York, p 12

Rogers EM (2010) Diffusion of innovations. Simon and Schuster, New York

Ryan B, Gross NC (1943) The diffusion of hybrid seed corn in two iowa communities. Rural Sociol 8(1):15

Saito K, Ohara K, Yamagishi Y, Kimura M, Motoda H (2011) Learning diffusion probability based on node attributes in social networks. In: Foundations of intelligent systems, Springer, pp 153–162

Sanchez S (2001) Abc's of output analysis. In: Proceedings of the 33nd conference on Winter simulation, vol 1, IEEE Computer Society,pp 30–38

Schreinemachers P, Berger T, Aune JB (2007) Simulating soil fertility and poverty dynamics in uganda: a bio-economic multi-agent systems approach. Ecol Econ 64(2):387–401

Schreinemachers P, Berger T, Sirijinda A, Praneetvatakul S (2009) The diffusion of greenhouse agriculture in Northern Thailand: combining econometrics and agent-based modeling. Can J Agric Econ Rev Can Agroecon 57(4):513–536

Schreinemachers P, Potchanasin C, Berger T, Roygrong S (2010) Agent-based modeling for ex ante assessment of tree crop innovations: litchis in Northern Thailand. Agric Econ 41(6):519–536

Schwarz N, Ernst A (2009) Agent-based modeling of the diffusion of environmental innovationsan empirical approach. Technol Forecast Soc Change 76(4):497–511

Schwoon M (2006) Simulating the adoption of fuel cell vehicles. J Evolut Econ 16(4):435–472

Shakarian P, Bhatnagar A, Aleali A, Guo R, Shaabani E (2015) Diffusion in social networks. doi:10.1007/978-3-319-23105-1

Sopha BM, Klöckner CA, Hertwich EG (2013) Adoption and diffusion of heating systems in norway: coupling agent-based modeling with empirical research. Environ Innov Soc Trans 8:42–61

Sorda G, Sunak Y, Madlener R (2013) An agent-based spatial simulation to evaluate the promotion of electricity from agricultural biogas plants in germany. Ecol Econ 89:43–60

Stonedahl F, Rand W (2014) When does simulated data match real data? Comparing model > Calibration Functions using Genetic Algorithms. Springer, Berlin

Stummer C, Kiesling E, Günther M, Vetschera R (2015) Innovation diffusion of repeat purchase products in a competitive market: an agent-based simulation approach. Eur J Oper Res 245(1):157–167

Sun Z, Müller D (2013) A framework for modeling payments for ecosystem services with agent-based models, bayesian belief networks and opinion dynamics models. Environ Model Softw 45:15–28

Train KE (2009) Discrete choice methods with simulation. Cambridge University Press, Cambridge

Tran M (2012) Agent-behaviour and network influence on energy innovation diffusion. Commun Nonlinear Sci Nume Simul 17(9):3682–3695

Trusov M, Rand W, Joshi YV (2013) Improving prelaunch diffusion forecasts: using synthetic networks as simulated priors. J Market Res 50(6):675–690

Vag A (2007) Simulating changing consumer preferences: a dynamic conjoint model. J Bus Res 60(8):904–911

Valente TW (1995) Network models of the diffusion of innovations. Hampton Press

Valente TW (2005) Network models and methods for studying the diffusion of innovations. Models Methods Social Netw Anal 28:98

Valente TW, Rogers EM (1995) The origins and development of the diffusion of innovations paradigm as an example of scientific growth. Sci Commun 16(3):242–273

Van Vliet O, De Vries B, Faaij A, Turkenburg W, Jager W (2010) Multi-agent simulation of adoption of alternative fuels. Transp Res Part D Transp Environ 15(6):326–342

Watts DJ, Strogatz SH (1998) Collective dynamics of small-worldnetworks. Nature 393(6684):440–442

Windrum P, Fagiolo G, Moneta A (2007) Empirical validation of agent-based models: alternatives and prospects. J Artif Soc Soc Simul 10(2):8

Wolf I, Nuss J, Schröder T, de Haan G (2012) The adoption of electric vehicles: an empirical agent-based model of attitude formation and change. In: Conference of the European association for social simulation, pp 93–98

Xiang X, Kennedy R, Madey G, Cabaniss S (2005) Verification and validation of agent-based scientific simulation models. In: Agent-directed simulation conference, pp 47–55

Zhang T, Gensler S, Garcia R (2011) A study of the diffusion of alternative fuel vehicles: an agent-based modeling approach. J Prod Innov Manag 28(2):152–168

Zhang H, Vorobeychik Y, Letchford J, Lakkaraju K (2016) Data-driven agent-based modeling, with application to rooftop solar adoption. Auton Agents Multi-agent Syst 30(6):1023–1049

Zhao J, Mazhari E, Celik N, Son YJ (2011) Hybrid agent-based simulation for policy evaluation of solar power generation systems. Simul Model Pract Theory 19:2189–2205