Thorben Jensen, Émile J.L. Chappin

Automating agent-based modeling

Data-driven generation and application of innovation diffusion models

Originally published in:
Environmental modelling & software,
92 (2017), 261-268
DOI: 10.1016/j.envsoft.2017.02.018

Wuppertal Institut

*Thorben Jensen a,b,**
Émile J. L. Chappin a,b

Automating agent-based modeling

Data-driven generation and application of innovation diffusion models

a Wuppertal Institute for Climate, Environment and Energy, Wuppertal, Germany

b Delft University of Technology, Delft, The Netherlands

* Corresponding author:
Thorben Jensen
Wuppertal Institute for Climate, Environment and Energy
Doeppersberg 19
42103 Wuppertal
Germany

E-mail: thorben.jensen@wupperinst.org

Wuppertal Institut

Automating agent-based modeling: data-driven generation and application of innovation diffusion models

Thorben Jensen[a,b,*], Émile J.L. Chappin[b,a]

[a] Wuppertal Institute for Climate, Environment and Energy, P.O. Box 100480, 42004 Wuppertal, Germany
[b] Delft University of Technology, P.O. Box 5015, 2600 GA Delft, The Netherlands

Abstract

Simulation modeling is useful to understand the mechanisms of the diffusion of innovations, which can be used for forecasting the future of innovations. This study aims to make the identification of such mechanisms less costly in time and labor. We present an approach that automates the generation of diffusion models by: (1) preprocessing of empirical data on the diffusion of a specific innovation, taken out by the user; (2) testing variations of agent-based models for their capability of explaining the data; (3) assessing interventions for their potential to influence the spreading of the innovation. We present a working software implementation of this procedure and apply it to the diffusion of water-saving showerheads. The presented procedure successfully generated simulation models that explained diffusion data. This progresses agent-based modeling methodologically by enabling detailed modeling at relative simplicity for users. This widens the circle of persons that can use simulation to shape innovation.

Keywords: Agent-based modeling, automated model generation, diffusion of innovations, data-analysis tool, policy simulation

1. Introduction

Understanding the prospects of innovations and how they spread is powerful. Mechanistic understanding of the diffusion of an innovation can help explaining their success. For instance, the Theory of Diffusion of Innovations by Rogers [1] allows understanding diffusions based on general mechanisms of interpersonal interactions. From these, it is possible to infer general patterns and key actors of innovation diffusion. Further, the explanatory power of the general mechanisms of innovation diffusion has been confirmed in empirical cases of diffusing innovations [2, 3].

Beyond understanding, found mechanisms can be used for guiding practical actions. Persons and organizations often want to know *"how to speed up the rate of diffusion of an innovation"* [1]. Actions that achieve this can directly be derived from causal mechanisms of the spreading of an innovation. Further, simulation can be used to project and estimate the impact of practical actions. This allows forecasting the impact of these actions from the underlying mechanisms. This paper will focus in particular on simulating innovation diffusion with agent-based modeling (ABM). This approach represents real-world actors with computer agents, whose actions towards innovations are modeled by explicit decision models [5, 6].

However, mechanistic understanding is particularly challenging to gain. It is harder to achieve than statistical inference, which reveals co-occurrence of events in a set of observations. Requirements for gaining it also exceed sole causal understanding, which 'only' requires knowing that one event generally causes another one [7]. Instead, mechanistic understanding implies to know if one event (likely) *"leads to a specific, deterministic behavior in another"* [8].

ABM can illuminate mechanisms of the diffusion of inno-

*Corresponding Author

Email addresses: `Thorben.Jensen@wupperinst.org` (Thorben Jensen), `E.J.L.Chappin@tudelft.nl` (Émile J.L. Chappin)

URL: `http://thorbenjensen.de` (Thorben Jensen), `http://chappin.com/emile` (Émile J.L. Chappin)

vations, but is challenged by time and labor intensive model building [9]. Via simulation, ABM links micro-level actions of actors to 'emergent dynamics', e.g. innovation diffusion [10]. Thereby, macro-dynamics of innovation diffusion are 'decoded' by being explained by micro behavior of agents [11, 12]. Unfortunately, ABM is commonly more time-intensive than its alternatives (e.g. system dynamics [13] and statistical analysis). This limits its practical applicability. In line with these challenges, Garcia and Jager [14] emphasize the current *"challenge of defining AMSs (i.e. agent-based simulation models) that are useful (to) managers without programming skills."*

We propose to enable agent-based modeling to overcome these limitations by automated model generation. Several approaches to automation exist, which we propose to combine: (1) Translating simple specifications into executable models. Examples are `http://m.modelling4all.org` and the MAIA framework [15], which automatically generate simulation models from specifications by domain-experts. (2) Model building from existing components. A methodology for this idea is 'TAPAS'[1], via which previously validated models are reused for new applications [16]. (3) Using data for model-building in a structured way. Grimm et al. [11] proposed 'Pattern-oriented Modeling' to falsify model variants that fail to reproduce a set of patterns from empirical data. This replaces ad-hoc decisions and informed guesses about adequate model structures and parameters with rigid testing against empirical data.

Therefore, the target of this study is to present a process that systematically builds ABMs via the following steps: (1) extracting driving mechanisms from empirical observations on innovation diffusion; (2) projecting diffusions into the future; and (3) assessing the effects of real-world actions and policies ex-ante, via simulation. This study aims to answer the following question: *"Can automated generation of agent-based models on the diffusion of innovation be achieved, and how could this be useful?"* This question will be addressed by specifying an automated software procedure for this task. To further provide *proof of concept*, application of an implementation of this procedure to the diffusion of sustainable products among households will be presented.

The remainder of this paper is structured as follows. First, we provide background on agent-based modeling of the diffusion of innovations. Second, the procedure that automates the building of such models is presented. Finally, this procedure is applied to a case of innovation diffusion.

2. Agent-based modeling of innovation diffusion

This section will provide details on agent-based modeling of innovation diffusion, which is the application domain of the proposed automation procedure. We will show that there exists a high degree of standardization of existing diffusion models. This standardization helps automated modeling.

According to Geels and Johnson [17], there exist four general types of dynamic innovation diffusion models. We hereby focus on innovation models that are dynamic, because innovation itself is a process of change [18]. (1) *Adoption models* capture spreading of an innovation among potential adopters, e.g. how the user base of a new product increases via word-of-mouth. (2) *Models of up-scaling and system building* describe a small system expanding to a larger one, e.g. an electricity system expanding from a decentralized ones to a single centralized system. (3) *Replication and circulation models* emphasize the replication of an adoption during its circulation to other location. Considering replication emphasizes adapting an innovation to other local conditions. (4) *Societal embedding models* consider the embedding of an innovation in business, societal, policy, and user environments.

'Adoption' type models are of special interest to this study. This is because their modeling of *"independent adopters making (adoption) decisions"* [17, p. 12] fits well with the actor-centric perspective of agent-based modeling. Adoption type models are represented by 'aggregated' and 'individual level' models [18]. Aggregated models directly model the overall adoption dynamics of an entire population. This approach is represented by the 'Bass model' and commonly modeled with system dynamics [18]. Conversely, 'individual level' models

[1] 'TAPAS' abbreviates *"Take A Previous model and Add Something"*.

capture the adoption decisions of individuals in a population, from which overall adoption dynamics ‘emerge’.

In this study, we will focus on the individual level models, because of their capability to incorporate more aspects of reality. According to Kiesling et al. [18], ‘individual level’ models are superior to ‘aggregated’ ones (such as system dynamics). (1) *Explanatory power* is greater for ‘individual level’ models, because they explicitly connect behavior and decisions of agents with aggregated diffusion dynamics. (2) *Population heterogeneity* can be captured more detailed in ‘individual level’ models. (3) *Social processes* (e.g. interactions between consumers) are modeled explicitly. This process can have great impact on diffusion success [5]. Agent-based ‘individual level’ models are particularly suited to model social interactions. In contrast to discrete-event simulation, they are capable of modeling detailed social interaction topologies in a computationally efficient way [13]. Consequently, this study will focus on innovation diffusion models that are agent-based.

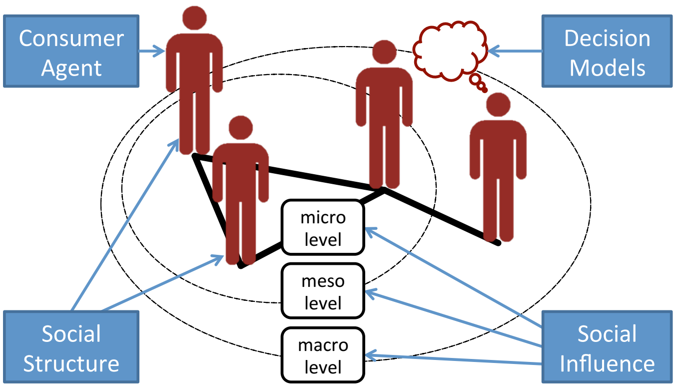

Fig. 1. Meta-model of agent-based models of innovation diffusion. Based on review by Kiesling et al. [18, Fig. 3].

Automating the building of agent-based innovation diffusion models is facilitated by their similar structure. A review by Kiesling et al. [18] finds that most ‘individual level’ diffusion models have such a common structure. Accordingly, virtually all agent-based innovation diffusion models are variations of one *meta-model*, shown in Fig. 1. This meta-model comprises the following elements: (1) Consumer agents represent the entities than can adopt an innovation. These can be individual persons, households, or groups of households. (2) Social structure is the heterogeneity of consumer agents, e.g. dividing them in different consumer groups. (3) Decision making processes (formalized as *decision models*) are the key actions of consumer agents to model the adoption of an innovation. (4) Social influence between agents (from peers, social groups or overall population) can affect decision making of consumers and is commonly modeled as a social network graph. This overall similarity simplifies automated model generation. This is because there is less variation in input data and less structural variation than needs to be considered.

3. Methods

In this section, we will present in detail the automation procedure to building agent-based models on innovation diffusion. We regard this approach as innovative, because it meets a previously unmet demand and was apparently not met this was previously. According to Garcia and Jarger [14], a *“versatile method of easily testing managerial strategies that influence the degree and speed of diffusion processes is not currently available.”* When querying the Scopus database for *‘agent-based AND innovation AND automat*’*, no existing similar approach was found.

The automation procedure will be presented by describing it conceptually and by giving details on its implementation.[2] Thereafter, proof of concept is given with an application case.

3.1. Automation procedure concept

We coin a method as specified in Fig. 2, comprising the three phases *preprocessing*, *inverse modeling*, and *policy simulation*.

Preprocessing. This phase is coined *preprocessing*, because input by the user is not given as raw data, but has to be preprocessed. The following types of input data are strictly required for the presented method to execute:

(1) Input data is provided on agents (i.e. the decision-making entities in an agent-based model). For each agent, its location

[2]Source code of the prototype implementation can be accessed at https://github.com/ThorbenJensen/automated-model-generation

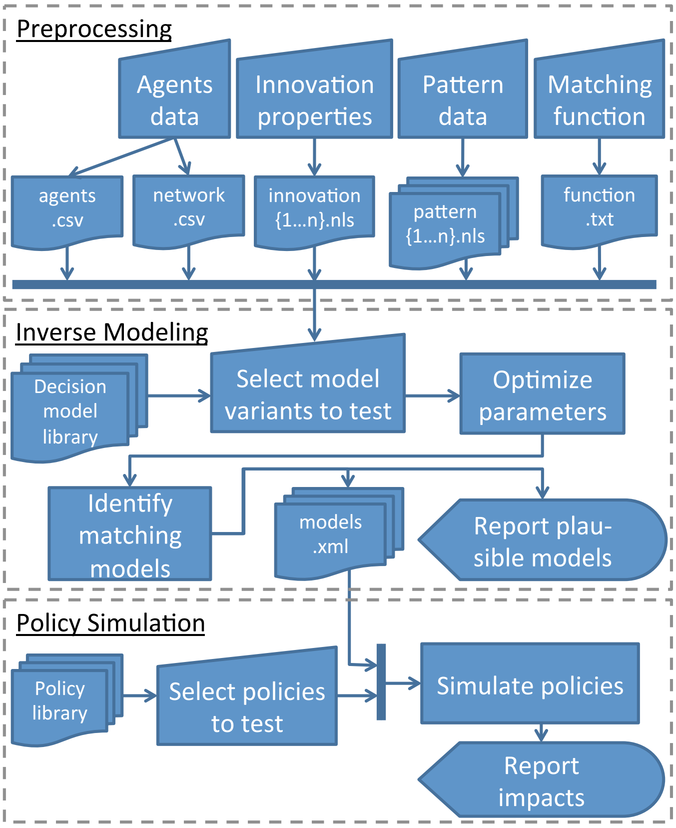

Fig. 2. Overview of phases of automation procedure. The procedure is sub-divided into the subsequent phases *preprocessing* of input data, *inverse modeling* of potentially explaining models, and *policy modeling* of models that were accepted based on the previous phase.

and social group are defined. This attribute of a social group enables us to capture the heterogeneity of agents. Social influence is defined by a social network graph. For generating a social network graph, we used the algorithm described by Jensen et al. [4, Appendix A.2].

Agents have to be defined by a CSV file with the columns ID, X and Y coordinates, and name of the social group they may belong to. The network graph is provided as a CSV file with the columns FROM and TO, defining directed links between two agents identified by their IDs. For instance, bidirectional influence between two agents would require two lines in this file.

(2) Innovation properties are provided that represent how an innovation is perceived by households. This idea follows Rogers [1], according to whom diffusion success of innovations depends on generalizable properties. Examples of the innovation properties are relative compatibility, complexity, and trialability.

Innovation properties each have to be provided as NetLogo source files. Each file contains a NetLogo method that sets innovation properties of an innovation as global variables.

(3) Patterns are provided that characterize the dynamics of the real-world process that shall be modeled. These patterns are *"indicators of essential underlying processes and structures"* [11]. Each additional pattern reduces uncertainty about which mechanisms could explain the diffusion of an innovation. An example for a relevant pattern is the exponentially increasing adoption share of a successful innovation during its initial diffusion [1].

Patterns are formalized by provided as NetLogo functions that calculate how well a simulation run matches each pattern. The values returned from these functions represent how well a simulation run suffices a pattern. A returned value of 0 signals a perfect fit with a pattern. With greater divergence from the pattern, this returned value increases. At simulation runtime, these functions query simulation runs and return fitness values for the following matching function.

(4) A 'matching function' describes the desired behavior of an accepted simulation model in terms of the provided patterns. This function weights and combines patterns to describe model output that would be considered realistic. This function assists in finding simulation runs that represent the empirical patterns best.

The matching function has to be defined by the user and passed as in a character sequence. Variables of this function are the names of the provided empirical patterns (and the functions that calculate matching with these patterns). For an example, see Eq. 2 at the application case below.

Inverse modeling. The inverse modeling phase identifies models that satisfy the provided matching function.

Within a range of plausibility, pre-defined models are varied in their structure and parameter values. For this, the NetLogo tool *BehaviorSearch* was used [20]. It repeatedly runs each potential model, thereby varying its structure and parameters, searching for an optimal fit with the pattern. The optimum that this search converges to is defined by the user-provided matching function. For the application case, we executed BehaviorSearch with a *simulated annealing* optimization (see Table 1 for search settings).

At the end of this phase, the user has to choose which tested models from the model library with which structural variation shall be accepted. Accepted model variants should be those that generate realistic results. This decision can be based on the best fitness values and respective parameters, which are reported for each structural variation of each tested model. If a model reproduced all provided empirical patterns, then it can be considered a potential explanation of these input data. Because the user has pre-defined this ideal behavior via the matching function, the fitness value is a strong indicator for this judgement. If model variants of multiple complexity levels match the patterns well, the simplest ones of these variants should be preferred. This serves to manage the risk of 'overfitting' at high structural complexity [19]. If required, the reported parameters settings for the best fit of each model variation allow the user to simulate and assess these model settings more closely in NetLogo.

Table 1

Search setting of simulated annealing optimization. Applied search tool was the NetLogo extension BehaviorSearch. Search parameters are names as in this tool.

Search parameter	**Value**
Mutation rate	0.05
Temperature change factor	0.95
Initial temperature	1.0
Restart after stall count	0
Evaluation limit	300
Optimization goal	'Minimize Fitness'
Collected measure	'MEDIAN_ACROSS_STEPS'
Fixed sampling	5
Combine replicates	'MEDIAN'

Policy simulation. The proposed automation procedure provides the useful function of semi-automatically assessing policies. Here, policies are those actions that aim at systematically supporting the diffusion of an innovation. Polices are provided in a policy library, which can be extended by the user. Such automated policy modeling is useful, first, because it frees the user from redundant, manual work. Further, running the same set of policies across all models that are accepted by the user based on the inverse modeling results increases robustness of the policy assessment. This can for instance be achieved by averaging over all these forecasts.

Policies are pre-implemented as NetLogo functions and stored as individual NetLogo source files. Users have to choose from a set of policies that support innovation diffusion or define other policy options. The user is recommended to test those policies for all diffusion models that resulted in a sufficient fit with the provided empirical patterns. Each policy simulation is executed from an XML file with the 'BehaviorSpace' tool in NetLogo. These files are derived from a template, but parsed based on the user choices on policies and models, and the respective parameterizations that previously resulted in a best match with the empirical data.

3.2. Application case: diffusion of water-saving appliances

We applied the here presented automation procedure to the diffusion of water-saving showerheads. This was motivated by available empirical data of high quality for this case. We used the proposed automation procedure to generate models that explain these data and to test policies. This served as a proof of concept and illustrates the proposed automation procedure. Also, it informs us about the mechanisms with which water-saving showerheads could spread. Policy simulation shows how this spreading could be effectively influenced.

3.2.1. Empirical data on application case

Empirical data on the diffusion of water-saving showerheads was used, as presented by Schwarz [21].

(1) *Agents data.* Previous research found a significant relationship between lifestyle group and adoption behavior regarding water-saving appliances [21]. Accordingly, three consumer groups could be clustered: 'Leading Lifestyles', which are of higher social status, are most interested in the adoption of such

appliances; ‘Mainstream and Traditional’ households show intermediate interest in them; and ‘Hedonists’ are least interested in water-saving appliances.

(2) *Innovation properties*. Properties of water-saving showerheads and conventional showerheads were surveyed. For each lifestyle group, the relative importance of these properties was also surveyed. This allows modeling the choice of consumers regarding the adoption of water-saving showerheads.

(3) *Diffusion patterns*. Two empirical patterns on the diffusion of water-saving showerheads emerged. First, marketing shares in Germany after 15 years of product diffusion show difference in adoption between these consumer groups. Second, the adoption diffusion curve during the first 15 years of innovation diffusion has an exponential shape.

3.2.2. Existing model on showerheads diffusion

An agent-based simulation model was previously built based on some of this empirical data [21]. We will here coin it the ‘Schwarz’ model. This model describes the decision making of agents regarding the adoption of feedback devices. According to the model, initially no household uses water-saving shower heads. At a monthly deliberation probability of 0.004, each household decides whether to adopt the water-saving option. There is a probability at which agents adopt the technology option that is adopted by the majority of their peers. This probability is differentiated by the three lifestyle groups [22]: (1) Leading Lifestyles always adopt the device, regardless of their peers; (2) Mainstream agents adopt devices in 50% of the cases, and imitate their peers otherwise; and (3) Hedonists always imitate the majority of their peers.

3.2.3. Evaluated agent-based models

We created a generic model library of two further models. We coined these models *‘Schwarz flexible’* and *‘TPB’*, which abbreviates Theory of Planned Behavior.

‘Schwarz flexible’ model. This model is structurally similar to the ‘Schwarz’ model, but its parameterization was made ‘flexible’ in two ways. First, the monthly deliberation probability became a flexible parameter between 0.004 and 0.04. Second, the probability of agents to adopt according to the majority of their peers also became a flexible parameter (between 0 and 1) for each social group.

‘Theory of Planned Behavior’ model. The second decision model is based on Ajzen’s [23] *Theory of Planned Behavior* (TPB). Modeled adoption is based on three factors: the *attitude* towards an innovation (ATT), the *perceived behavioral control* (PBC) over adopting it, and the *subjective norm* (SN) towards adoption from the social environment. For water-saving showerheads, this means that adoption is more likely if first, attitude towards this product is more positive, second, if the adoption is perceived as easy and feasible, and third, if adoption is more common among peers. We used the formalization shown in Eq. 1 [21].

$$\text{adoption\_intention}_i = (1 - s) \cdot (\text{ATT}_i + \text{PBC}_i) + s \cdot \text{SN}_i \quad (1)$$

According to this model, an agent calculates utility for each option i and adopts the one with the highest adoption intention, based on the following factors. ‘ATT_i’ is the product of two vectors: properties of innovation i and weights (i.e. importance) that the agent’s social group assigns to these characteristics. An example of such a characteristic is environmental-friendliness of an innovation. ‘PBC_i’ is a product of innovation characteristics (that translate into the ease of adoption) and the respective weights of importance for the social group. An example is the purchasing cost. ‘SN_i’ is the ratio of peers of a household that use product *‘i’*. The parameter *‘s’* is the importance to practice the same behavior as its peers, motivated by a need for social cohesion or uncertainty about the product.

We differentiated these two models by an optional word-of-mouth (WOM) mechanism. Without this mechanism being active, all agents can principally deliberate on adoption at any time. If this mechanism is active, agents only consider adopting feedback devices if they are *aware* of them. At adoption, an agent makes the peers that it influences aware of the device. The activation of this mechanism thus becomes an additional degree of freedom to the structure of both models. In the inverse modeling phase of the automation procedure, this will become subject to structural model variation.

3.2.4. Automated policy simulation

In addition to enhancing mechanistic understanding, we assessed the impact of policy actions towards innovation diffusion. A policy (i.e. *"course or principle of action"* [24]) regarding innovations often aims at directing their diffusion [6]. Typically, this is increasing their rate of diffusion.

The above presented automation procedure can automatically project the impact of policies on diffusion. This could be used to test implementations of new policies, as well as the termination of previous ones. The automation phase only uses those models for projections of policy impacts that were accepted based on the inverse modeling phase.

As policies to be tested, we chose two marketing strategies at which free products are given away. (1) After 15 years of device diffusion, an additional 10% of households receive a free water-saving shower head. (2) The same policy is applied, but to those households who influence most other households. These selected households can be framed as households of *opinion leaders*, who are highly connected and influential [18]. They have thus shown particular potential to leverage innovation diffusion [1, 18, 25, 26]. Simulation of this second policy relies on the explicit modeling of the social network. Consequently, it could not directly be tested by some simulation approaches that lack a modeled social network, e.g. system dynamics.

The tested policies have the potential to promote further adoption of this product by social influence and WOM. Time of policy implementation is 15 years after the beginning of product diffusion. From this point in time, no empirical data were available. Policy simulation thus projects the uncertain future diffusion.

4. Results and discussion

We conducted two simulation experiments, each representing one of the two automated phases of the procedure.

- Experiment 1 simulates the simulation models from the model library and compares simulation results to the original 'Schwarz' model.
- Experiment 2 demonstrates automated policy simulation with the models that were accepted as sufficiently realistic in the first experiment.

4.1. Experiment 1: Inverse modeling

In this experiment, two diffusion models ('Schwarz flexible' and 'Schwarz TPB') were tested for their ability to explain the historical diffusion of water-saving showerheads. This testing is taken out by the inverse modeling phase of the proposed automation procedure. Each of these two models was simulated at two structural variations (with and without the WOM mechanism) and at varied parameters. Simulation results were tested against two empirical patterns: the exponential takeoff of adoption and the empirical market shares of the three consumer groups after 15 years.

The provided matching function that was *minimized* in order to search for realistic models is shown in Eq. 2. Mainly, the simulated adoption shares are compared to the provided empirical ones. In the inverse modeling phase, mismatching with empirical market shares is minimized. Further, if the shape of the adoption curve is not exponential, then a significant penalty is added to the matching function. Basis for this is the overall adoption share over all agents and the length of a simulation run of 15 years. Matching results (i.e. best fitness and according parameters) are shown in Table 2.

$$\textbf{minimize}\ \{\text{`adoption shares'} + 1000 \cdot \text{`exponential'}\} \qquad (2)$$

Results of best matches, shown in Fig. 3, revealed that model versions without WOM were less able to match the patterns: the Schwarz flexible model, was not able to generate an exponential pattern, while the TPB model could generate exponential increase in adoption, but was not able to match the adoption data at the same time. With the WOM mechanism being active, both models were able to match both patterns. The only

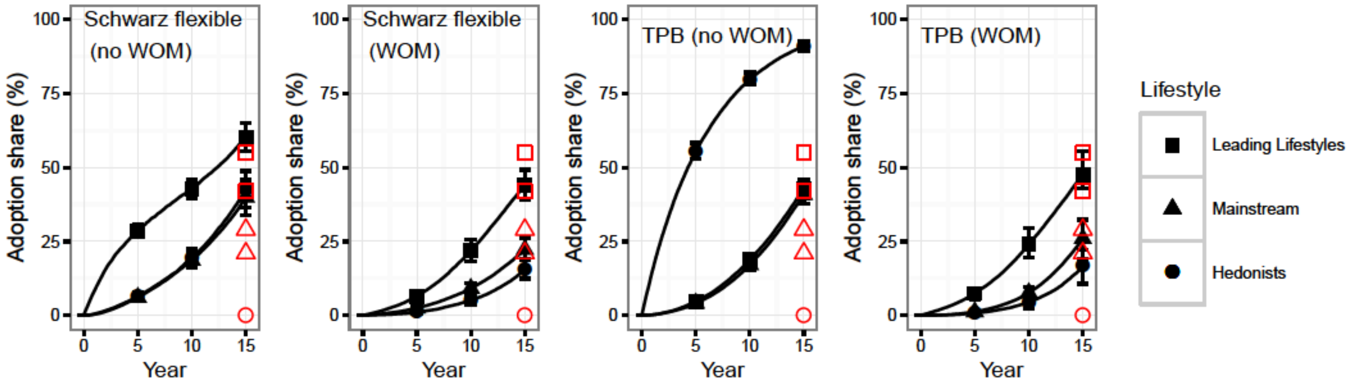

Fig. 3. Average adoption of water-saving showerheads, as simulated by the four tested model structured at best matching parameters. Results are differentiated by consumer group. Whiskers show the quartiles. The hollow points show empirical market shares of the respective consumer group after 15 years of diffusion. For each of the consumer groups Leading Lifestyle and Mainstream, two market share data points were used.

Table 2

Results of inverse modeling phase: best fit and parameterizations. Optimized fitness for the models 'Schwarz flexible' and 'TPB' with and without word-of-mouth (WOM) is shown. Parameter combinations (except those that resulted in no adoption at all) with best fit are shown: the monthly deliberation probability and social influence (δ_α) in adoption are given for the consumer group 'Leading Lifestyles', 'Mainstream and Traditionals', and 'Hedonists' (s_{LL}, s_{MS}, s_{HD}).

Model	WOM	fitness	δ_α	s_{LL}	s_{MS}	s_{HD}
'Schwarz'	no	-	0.004	0	0.5	1
'Schwarz flex.'	no	19.12	0.029	0.723	1	0.996
'Schwarz flex.'	yes	5.91	0.013	0	0.679	0.928
'TPB'	no	26.61	0.013	0.288	0.428	0
'TPB'	yes	5.72	0.016	0	0.456	0.200

limitation to this matching is a relatively bad reproduction of the empirical market share of the Hedonists group. Based on these results, we regard both simulated models generally suited to explain the diffusion of water-saving showerheads, but only if the WOM mechanism is included.

4.2. Experiment 2: Policy simulation

In this experiment, we applied the proposed procedure to automatically assess the impact of a policy on innovation diffusion. This assessment only based on those model variants that matched the empirical patterns in the previous experiment. Instead of testing policy interventions for one simulation model, policies are tested for all models that were thus accepted in the inverse modeling phase. The simulated policies (see Section 3.2.4) are as follows: (1) to give away free water-saving showerheads to 10% of households after 15 years of innovation diffusion; and (2) giving away water-saving showerheads at the same point in time to 10% of households, who are influencing the most other households (i.e. who have outgoing network connections to most other households).

Figure 4 and 5 show the impact of the assessed policies, which led to the following findings. First, impacts for the two models are relatively similar: giving away free devices at the advanced stage of product diffusion makes the scenarios with and without policy intervention initially diverge quickly. Following the interventions, the innovation spreads at a similar rate, compared to the reference scenario without intervention. Second, for both models, the higher adoption due to the intervention led to a gradual saturation in adoption at the end of 25 years of diffusion. Adoption over time thus forms an S-curve, which is predicted by the Theory of Diffusion of Innovations [1]. This shows that (in this regard), the simulated models are in line with prevailing theory. Overall, the similar additional impact for the two models underlines the robustness of the proposed procedure.

The two assessed policies had a different impact. For both used models, addressing opinion leaders generated a higher impact than addressing random households. Further, the similarity in policy impact for the two simulated models and the difference between the policies is underlined in Table 3. It shows

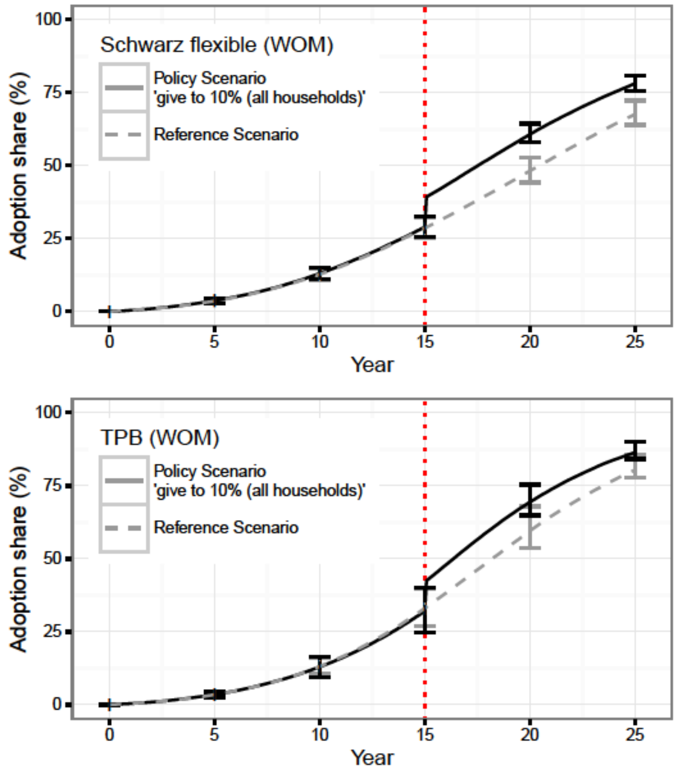

Fig. 4. Impacts of policy that addresses all households (continuous line) compared to baseline scenario (dashed line). Whiskers show the quartiles. Results rely the two most realistic model structures with parameterizations that matched empirical patterns best.

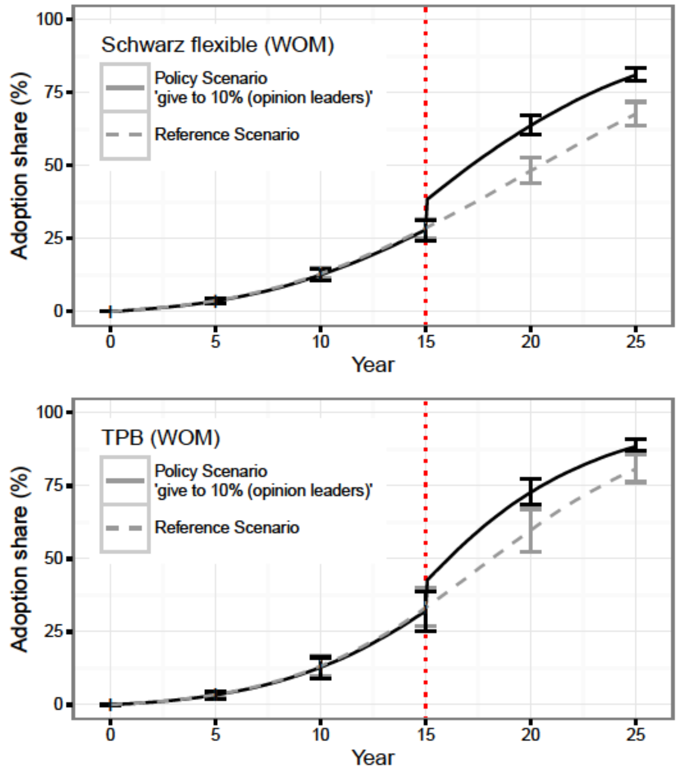

Fig. 5. Impacts of policy that addresses opinion leaders (continuous line) compared to baseline scenario (dashed line). Whiskers show the quartiles. Results rely the two most realistic model structures with parameterizations that matched empirical patterns best.

the same relative order of impact of the two assessed policies. For both models, the marketing strategy of addressing opinion leaders has a higher impact. Further, the impact of each policy (compared between both models it was tested with) is relatively similar. At this point, it would be possible to extract statistical properties of predicted policy impacts over all tested models. For estimating the expected impact, averaging of predictions would be advisable. Alternatively, minimum and maximum of such an ensemble would give insights into degree of uncertainty. Overall, this indicates that the policy assessment based on multiple models increased the robustness of the proposed procedure.

4.3. Limitations

Discussion of limitations will focus on two aspects of the proposed automation procedure rather than the application case. This is because this procedure is the key contribution of this study.

(1) The proposed automation procedure might not be applicable to very uncertain processes or models. It appears limited to cases where potential explanations are restricted to a bounded space of options. This is the case for e.g. innovation diffusion. Nevertheless, the proposed procedure has been able to handle structural uncertainty. However, up to which limit such uncertainty can be managed is not known at this point.

(2) The proposed procedure is not easily applicable by everyone. It requires data processing skills in the preprocessing phase. This might limit the circle of potential users. Yet, the procedure still widens this circle of users, compared to the prevailing model building 'from scratch'.

(3) Further, the procedure might require cautious application by the user. Even though the presented method is mostly automated, key decisions still have to be made by the user. This critical role of user decisions is a common feature of automated

Table 3

Results of policy simulations based on selected, sufficiently realistic models (with word-of-mouth). Impact is shown as additional percentage of product adoption 15 years after policy implementation.

Model	WOM	Policy	Add. adoption (10 yrs)
'Schwarz flex.'	yes	*'give away to 10%'*	10.5%
'Schwarz flex.'	yes	*'give away to 10% (opinion leaders)'*	13.3%
'TPB'	yes	*'give away to 10%'*	6.0%
'TPB'	yes	*'give away to 10% (opinion leaders)'*	7.7%

data-analysis tools, e.g. statistical tests [27]. If these decisions are not cautiously made in the presented automation procedure, quality of results might be compromised. For instance, tested diffusion models might be selected by the user without understanding their functioning.

5. Conclusion

The question guiding this study has been how the generation of agent-based innovation diffusion models can be automated and how this could be useful. This question has been addressed by specifying and presenting an automation procedure to the generation of agent-based models on innovation diffusion and by applying to a case study.

Implementation and application of the proposed design showed that the automation procedure is applicable to the diffusion of water-saving showerheads. It further enabled high efficiency of time and labor for this case. This serves as a proof of concept and adds weight of evidence to its suitability to automate the generation of agent-based models of innovation diffusion.

This application further revealed several advantages of the proposed automation procedure. Present practices of building agent-based models on innovation diffusion are highly diverse. Therefore, it does not seem informative to compare the here proposed procedure against any specific existing practice. Instead, we will conclude on the presented method by re-iterating its advantages. We stress that, in combination, these benefits validate the proposed design.

At application, the procedure proved helpful for improving existing diffusion models from empirical data. The previously empirically validated 'Schwarz model' on the diffusion of water-saving showerheads could be refined to increase its realism. For this refinement, word-of-mouth mechanism of communication between consumers was found plausible—both theoretically and data-wise. This role of word-of-mouth adds weight of evidence to the importance of future marketing efforts that leverage this mechanism.

The rigid use of data in the proposed procedure creates model validation by design. The procedure is driven by comparing model output to empirical data, which is central to validation [28]. Further, systematically comparing multiple models (and mechanisms) enables the good scientific practice of being able to falsify those that can not explain empirical observations. Overall, this has the potential to make agent-based modeling more rigorous than in common practice [11].

The presented approach allows using relatively complex simulation modeling at low complicatedness for the user. Provided a library of potential mechanisms has previously been implemented, a user would only need to provide key data on a dynamic, potentially complex system. The automated procedure then simulates bottom-up models and then tests their matching with the provided data. This procedure selects potentially explaining mechanisms and thus supporting gaining mechanistic understanding.

Due to this relative ease of use, the presented automation approach helps increasing the circle of persons that could independently build agent-based simulation models on innovation diffusion. We see the classical role of the modeler extended by the role of the *user* (also referred to as 'thematician' [29, 30]). Such a user can build and apply diffusion models without requiring programming or simulation skills. Except for extending a library of model components, the commonly required implementation by modelers and computer scientists [29, 30] is not required. A user only has to process and provide the required input data, as well as interpret the generated model results. From a perspective of innovation diffusion, we regard this widening of the circle of adopters a crucial service to the spreading of agent-based modeling as an innovative forecasting method.

5.1. Future research

We suggest to progress this study in three directions.

First, the central phase of inverse modeling is crucial to the proposed automation procedure and could be improved. We propose to support anticipated users of this automation procedure to make good choices on matching functions. For this, different designs of the inverse modeling phase should be compared. Those that are robust in providing good results over several applications cases should be preferred. One such variation would be to withhold for validation some of the data that is now used for model calibration. For choosing between alternating model hypothesis, various statistical approaches should be tested. Candidate methodologies for this are, for instance, Akaike Information Criterion and Bayes factors.

Second, user-friendliness of the procedure can be increased by accepting unstructured input data. The presented application case used structured empirical data. Approaches from data science could allow us to execute the procedure with un-structured data. Overall, increased user-friendliness further increases the circle of potential users.

Finally, we suggest to expand the application of the proposed automation procedure to more cases. This could be facilitated by finding a way for the automation procedure to be as generally applicable as possible. For instance, this could even include generating models from far smaller components than are currently in the modeling library. Application to more cases would eventually help establish *reference models* on the diffusion of innovations, which can further support the development of sound innovation diffusion models.

Overall, we believe these future development and applications will encourage users who are not model builders to apply the proposed automation approach. The here presented design is meant to assist them in exploiting the merits of agent-based modeling of innovation diffusion.

Acknowledgements

We thank our colleagues from the research project *EnerTransRuhr* (supported by the German Federal Ministry of Education and Research, grant number 01UN1205A) at Wuppertal Institute for providing empirical data. We further thank Prof. Paulien M. Herder, Prof. Klaus Troitzsch, Dr. Georg Holtz and Jochem Douw for comments on the draft of this paper. Finally, we would like to thank anonymous reviewers and audience of the *iEMSs 2016* conference for useful feedback on a previous version of this paper.

References

[1] E. M. Rogers, Diffusion of Innovation, 5th Edition, Free Press, New York, 2003.

[2] N. Schwarz, A. Ernst, Agent-based modeling of the diffusion of environmental innovations - an empirical approach, Technological Forecasting and Social Change 76 (4) (2009) 497–511. doi:10.1016/j.techfore.2008.03.024.

[3] B. M. Sopha, C. A. Klckner, E. G. Hertwich, Adoption and diffusion of heating systems in Norway: Coupling agent-based modeling with empirical research, Environmental Innovation and Societal Transitions 8 (2013) 42–61. doi:10.1016/j.eist.2013.06.001.
URL http://www.sciencedirect.com/science/article/pii/S2210422413000427

[4] T. Jensen, G. Holtz, C. Baedeker, E. Chappin, Energy-efficiency impacts of an air-quality feedback device in residential buildings: an agent-based modeling assessment, Energy and Buildings 116 (2016) 105–119.

[5] S. A. Delre, W. Jager, T. H. A. Bijmolt, M. A. Janssen, Targeting and timing promotional activities: An agent-based model for the takeoff of new products, Journal of Business Research 60 (8) (2007) 826–835. doi:10.1016/j.jbusres.2007.02.002.
URL http://www.sciencedirect.com/science/article/pii/S0148296307000392

[6] T. Jensen, E. Chappin, Changing domestic heating behavior: simulating marketing strategies for behavior-changing feedback devices, Technological Forecasting and Social ChangeUnder review.

[7] O. O. Aalen, A. Frigessi, What can statistics contribute to a causal understanding?, Scandinavian Journal of Statistics 34 (1) (2007) 155–168.

[8] J. T. Leek, R. D. Peng, What is the question?, Science 347 (6228) (2015) 1314–1315.

[9] K. H. van Dam, I. Nikolic, Z. Lukszo, Agent-Based Modelling of Socio-Technical Systems, Springer, Guildford and UK, 2012.

[10] E. J. Chappin, G. P. Dijkema, Modeling for transition management, Available at SSRN 2618413.

[11] V. Grimm, E. Revilla, U. Berger, F. Jeltsch, W. M. Mooij, S. F. Railsback, H.-H. Thulke, J. Weiner, T. Wiegand, D. L. DeAngelis, Pattern-oriented modeling of agent-based complex systems: lessons from ecology, Science 310 (5750) (2005) 987–991.

[12] N. Stern, Economics: Current climate models are grossly misleading., Nature 530 (7591) (2016) 407.

[13] C. Watts, N. Gilbert, Simulating Innovation: Computer-based Tools for Rethinking Innovation, Edward Elgar Publishing, 2014.

[14] R. Garcia, W. Jager, From the special issue editors: Agent-based modeling of innovation diffusion*, Journal of Product Innovation Management 28 (2) (2011) 148–151.

[15] A. Ghorbani, P. Bots, V. Dignum, G. Dijkema, Maia: a framework for developing agent-based social simulations, Journal of Artificial Societies and Social Simulation 16 (2) (2013) 9.
URL `http://jasss.soc.surrey.ac.uk/16/2/9.html`

[16] K. Frenken, Technological innovation and complexity theory, Economics of Innovation and New Technology 15 (2) (2006) 137–155.

[17] V. Geels, Frank W. Johnson, Adoption, upscaling, replication/circulation, and societal embedding: Four theoretical models of technology diffusion applied to biomass district heating systems in austria (1979-2013), in: Proc. International Sustainability Conference (IST, Helsinki, Finland, 2015), 2015.

[18] E. Kiesling, M. Günther, C. Stummer, L. M. Wakolbinger, Agent-based simulation of innovation diffusion: a review, Central European Journal of Operations Research 20 (2) (2012) 183–230.

[19] F. Provost, T. Fawcett, Data Science for Business: What you need to know about data mining and data-analytic thinking, " O'Reilly Media, Inc.", 2013.

[20] F. Stonedahl, U. Wilensky, BehaviorSearch (accessed online: 2016-04-21).
URL `http://behaviorsearch.org`

[21] N. Schwarz, Umweltinnovationen und lebensstile: eine raumbezogene, empirisch fundierte multi-agenten-simulation, Ph.D. thesis, Metropolis-Verl., Marburg (2007).

[22] T. Jensen, G. Holtz, E. J. L. Chappin, Agent-based assessment framework for behavior-changing feedback devices: spreading of devices and heating behavior, Technological Forecasting and Social Change 98 (2015) 105–119. `doi:10.1016/j.techfore.2015.06.006`.

[23] I. Ajzen, The Theory of Planned Behavior, Organizational Behavior and Human Decision Processes 50 (2) (1991) 179–211. `doi:10.1016/0749-5978(91)90020-T`.
URL `http://www.sciencedirect.com/science/article/pii/074959789190020T`

[24] O. U. Press, Policy, (accessed online: 2016-04-22).
URL `http://www.oxforddictionaries.com/definition/english/policy`

[25] M. C. Nisbet, J. E. Kotcher, A two-step flow of influence? opinion-leader campaigns on climate change, Science Communication.

[26] P. Van Eck, W. Jager, P. Leeflang, Opinion Leaders' role in innovation diffusion: A simulation study, Journal of Product Innovation Management 28 (2) (2011) 187–203. `doi:10.1111/j.1540-5885.2011.00791.x`.

[27] A. M. Strasak, Q. Zaman, K. P. Pfeiffer, G. Gobel, H. Ulmer, Statistical errors in medical research-a review of common pitfalls, Swiss medical weekly 137 (3/4) (2007) 44.

[28] W. Rand, R. Rust, Agent-based modeling in marketing: Guidelines for rigor, International Journal of Research in Marketing 28 (3) (2011) 181–193. `doi:10.1016/j.ijresmar.2011.04.002`.

[29] A. Drogoul, D. Vanbergue, T. Meurisse, Multi-agent based simulation: Where are the agents?, in: Multi-agent-based simulation II, Springer, 2002, pp. 1–15.

[30] J. M. Galán, L. R. Izquierdo, S. S. Izquierdo, J. I. Santos, R. Del Olmo, A. López-Paredes, B. Edmonds, Errors and artefacts in agent-based modelling, Journal of Artificial Societies and Social Simulation 12 (1) (2009) 1.