Energy Informatics

Reinforcement learning in local energy markets

Check for updates

Samrat Bose[1*] (iD), Enrique Kremers[1], Esther Marie Mengelkamp[2], Jan Eberbach[1] and Christof Weinhardt[3]

*Correspondence: bose@eifer.org
[1]European Institute for Energy Research, Emmy-Noether-Str. 11, 76131 Karlsruhe, Germany
Full list of author information is available at the end of the article

Abstract

Local energy markets (LEMs) are well suited to address the challenges of the European energy transition movement. They incite investments in renewable energy sources (RES), can improve the integration of RES into the energy system, and empower local communities. However, as electricity is a low involvement good, residential households have neither the expertise nor do they want to put in the time and effort to trade themselves on their own on short-term LEMs. Thus, machine learning algorithms are proposed to take over the bidding for households under realistic market information. We simulate a LEM on a 15 min merit-order market mechanism and deploy reinforcement learning as strategic learning for the agents. In a multi-agent simulation of 100 households including PV, micro-cogeneration, and demand shifting appliances, we show how participants in a LEM can achieve a self-sufficiency of up to 30% with trading and 41,4% with trading and demand response (DR) through an installation of only 5kWp PV panels in 45% of the households under affordable energy prices. A sensitivity analysis shows how the results differ according to the share of renewable generation and degree of demand flexibility.

Keywords: Agent-based simulation model, Bidding Strategies, Peer-to-peer trading, Local Energy Market, Reinforcement Learning, Demand Response

Introduction

The recent development of emerging technologies in the power industry has led to a paradigm shift in the frameworks and business models of the electricity retail market of the future (Chen et al. 2018). In 2017, the investment in renewable energy sources (RES) rose to 298 billion USD and it continues to increase with Europe having a share of 55 billion USD (International Energy Agency 2018). On May 2019, the European Commission (EC) has adopted the final files on Clean energy package for all Europeans which was placed in late 2016. The clean energy package contains the adoption of two directives with relevance to LEMs, including the Internal Electricity Market Directive (EU) 2019/944 which introduced the "Citizen Energy Community" and the Renewable Energy Directive (EU) 2018/2001 which introduced the "Renewable Energy Community" (Caramizaru and Uihlein 2020). These regulations describe the role of consumer participation in achieving

the flexibility which is essential to accommodate the variable and distributed renewable electricity generation in the electricity system .

The active engagement of end-users of electricity with the EC's target to make electricity 40% reduction in green house gas emissions by 2030 has paved the way for systematic incorporation of decentralised RES into the electricity system, and local energy markets (LEMs) provide a perfect platform for the entire ecosystem (Mendes et al. 2018). LEMs are targeted towards establishing a balance between the local generation and consumption which may facilitate a reduction in energy transmission, network congestion and expedite proper inclusion of decentralised RES (Mengelkamp et al. 2018a).

A robust LEM can be established through a well organised market mechanism. So, trading in the LEM is a vibrant topic of interest among the research communities, industry and policymakers (Mengelkamp et al. 2018a). The energy modelling community worldwide is focused on developing new trading approaches to replicate the decision-making process of the participants of the LEMs. The recent developments in the field of machine learning are providing answers to this research topic. Chen and Su (2018a) and Pilz and Al-Fagih (2017) have demonstrated the application of Q-learning and game theory approaches towards the development of trading strategy for LEMs. In spite of that, there is substantial research gap in this topic because very less literature is available developing trading strategies of residential prosumers. So, through this paper, we bridge the gap by demonstrating the application of reinforcement learning in building a trading strategy for participants of a residential LEM facilitated by DR.

Definitions and related work

Machine learning is a subset of artificial intelligence in the field of computer science that deals with certain algorithms and statistical models that machines use to perform a particular work through recognizing patterns and inferences instead of direct instructions from the user (Bishop 2006; Koza et al. 1996).

There are three branches of machine learning (Silver 2015):

1.   Supervised learning
2.   Unsupervised learning
3.   Reinforcement learning

Supervised learning approach or "learning with a teacher" is learning from a training set of labelled examples provided by a knowledgeable external supervisor "a teacher". It is called supervised because of the presence of the outcome variable to guide the learning process (Sutton and Barto 1998). In the unsupervised learning approach or "learning without a teacher", output data is given without any inputs. The goal is to discover interesting structures, associations or patterns in the data (Hastie et al. 2009). Situated in between supervised learning and unsupervised learning is the paradigm of reward (reinforcement) learning (RL). RL deals with learning in sequential decision-making problems in which there is limited feedback (Kaelbling et al. 1996). In RL, there is no supervisor, only a reward signal or a real number that tells the agent how good or bad was its action (Panait and Luke 2005).

We want to build intelligent agents who initiate human behaviour while trading. So, Modified Erev-Roth algorithm (Nicolaisen et al. 2001) under the reinforcement learning is chosen as a method of learning for the agent-based LEM here in our case because it is

the closest to the replication of the human decision-making process which is also established by psychological research (Mengelkamp et al. 2018c). In addition, there has been substantial research published on this topic and this particular algorithm is used mostly as a learning strategy in agent-based simulations in the energy sector (Mengelkamp et al. 2018c). So, using this algorithm will provide us a benchmark to test and analyse our results in comparison to the existing work.

Reinforcement learning

Reinforcement learning refers to the development of certain strategies, which software agents implement in order to learn how to maximize a certain cumulative reward through trial and error interaction with a dynamic environment (Kaelbling et al. 1996). The application of reinforcement learning for making financial decisions is demonstrated in Moody and Saffell (2001) and Maringer and Ramtohul (2012). Shimokawa et al. (2009) demonstrates the creation of an augmented learning model used to predict human behaviour while performing a financial investment task. Reinforcement learning has found its special application to optimize and automate the bidding strategies in different markets. Bidding strategy optimization in electricity markets through reinforcement learning is demonstrated in Wu and Guo (2004). A day ahead market model is empowered with reinforcement learning to assess the market power for various participants under auction-based energy pricing in Nanduri and Das (2007). Guo et al. (2009) have demonstrated through a multi-agent based model, how reinforcement learning can be applied at the appliance level for demand side management since only price-based constraints can negatively impact the system stability. Similar work for demand side management through binary control devices facilitated by reinforcement learning is done by Claessens et al. (2012). Claessens et al. (2013) exhibits a multi-agent-based system for demand response (DR) of a heterogeneous cluster of residential flexibility carriers. The results demonstrate that reinforcement learning is effective in peak shaving and valley filling with a faster convergence time.

Reinforcement learning in LEM

LEMs are defined as a group of electricity producers, prosumers and consumers who share the decentralised electricity produced among each other through an established trading mechanism in a closed geographical construct or a virtual community (Mengelkamp et al. 2018a). LEMs provide a powerful solution for energy decentralization along with several other benefits like enhancing the financial benefits for the agents of the community, ameliorating energy self-sufficiency of a community, or promoting local renewable energy generation (Koirala et al. 2016; Mengelkamp et al. 2018b; Olivella-Rosell et al. 2018). The application of reinforcement learning for microgrids through a multi-agent model is showcased in Dimeas and Hatziargyriou (2010). The model demonstrates the working of a microgrid in island mode operation. A similar approach for battery scheduling through reinforcement learning is applied in Kuznetsova et al. (2013) for an intelligent energy management system of a microgrid. The role of emerging brokers in a LEM at the distribution level to facilitate peer-to-peer energy trading is demonstrated in Chen and Su (2018a).

Automation of the bidding strategies relies on the structure of intelligent agents. Weidlich and Veit (2008) have given a survey of different categories of intelligent agent

strategies based on agent-based simulation models of wholesale electricity trading. The article showcases that the Erev-Roth algorithm with its modification is being used by a significant number of models. It is verified and established by further investigation in Mengelkamp et al. (2018c) who concluded the various reasons for the adoption of Erev-Roth learning mechanism over other intelligent agent strategies. The Erev-Roth algorithm (Erev and Roth 1998) modified by Nicolaisen et al. (2001) is able to imitate human learning behaviour which is one of the most important reasons behind using this algorithm for simulation of the learning behaviour of intelligent agents in energy markets. In this article in "Pricing strategy" section, we have explained how reinforcement learning algorithm is applied on a LEM to create the pricing strategy of the model.

Reinforcement learning application in LEM facilitated by DR

Demand side management (DSM) refers to all the measures taken on the energy consumption side to improve the efficiency of consumption. There are various methods of demand side management as analyzed in Palensky and Dietrich (2011) which includes energy efficiency (EE), time-of-use tariff (TOU), demand response (DR) and spinning reserve (SR). In this paper, we will discuss only DR. Albadi and El-Saadany (2008) have defined DR as a collection of all measures taken to modify consumption patterns in response to dynamic change in energy prices. It includes three major methods for load mediation i.e. load shifting to future time at favorable energy pricing time periods, local-generation, and load curtailment (Siano 2014). In this paper, we concentrate on the first two methods of local generation and load shifting to future time periods. In a real-world scenario, this demand shifting is realised through the use of smart devices, intelligent energy management system and user behaviour (Mengelkamp et al. 2018a; Jensen et al. 2018). This paper assumes that enough smart devices are available to sustain flexibility bids on the market. We point out that many households are currently not at this technological development stage, so that our paper and the subsequent market model will currently apply firstly to pioneer households with adequate flexibility providing smart devices. However, as the distribution of smart devices will increase, the number of potential households with adequate flexibility means will increase in time.

   DR demonstrates several advantages, which include maximizing the efficiency of renewable energy systems through load shifting from lower local generation times to higher local generation times, and optimize the required peak power installation of renewable energy systems through peak curtailment, which improves its cost-effectiveness (Mengelkamp et al. 2018a). DR also reduces the consumption cost of electricity through load shifting towards low energy price time periods (Albadi and El-Saadany 2008; Pinson et al. 2014). However, DR also has certain disadvantages, which hinders its full-scale application. One of the biggest barriers in DR application in Germany is the regulation, which does not yet provide profitable platform for residential DR applications in the current energy system of Germany (Mengelkamp et al. 2018a).

   Mengelkamp et al. (2018a) have listed numerous applications of DR. The application of smart grid information technology in empowering customers to participate in DR is demonstrated by Shariatzadeh et al. (2015). However, no model and quantifying results have been proposed to examine the efficiency of the mentioned strategy. Residential DR pilot projects have been modelled and analyzed worldwide. The pilot project of 40

Norwegian households with price-based DR is exhibited by Saele and Grande in Saele and Grande (2011).

Although, there is abundant literature available for DR, still, the application of DR in LEMs is explored by only a few researchers. Marzband et al. (2013) and Marzband et al. (2014) evaluated an energy management system in an island mode operation of a physical microgrid. The paper proposes a strategy based on gravitation search algorithm to solve the problem of DR. Mazidi et al. (2014) modeled the integrated scheduling of renewable generation and DR programs in a microgrid through forecasting of wind and solar irradiation for the day ahead energy market.

Research gap

Chen and Su (2018b) and Chen and Su (2018a) have explored the application of modified Q-learning algorithm for defining the trading strategies in a LEM. However, the results show that the modified Q-learning strategy proposed is only beneficial when the strategy is applied for long term so that the algorithm has sufficient time to learn. Mengelkamp et al. (2018c) presented a modification of modified Erev-Roth algorithm, which increases the self-consumption of the LEM by 15%. But the premise of either flexible generation or flexible DR was not investigated in the paper. Further, the increase of the size of the generator also influences the trading behaviour and benefits of the LEM, which is not explored in Mengelkamp et al. (2018a). Vázquez-Canteli and Nagy (2019) gives a review of algorithms and model techniques involving single and multi-agents presented by various researchers for application of reinforcement learning for DR. However, none of these papers investigates the impact of DR on trading behaviour of the participants in a LEM.

In this paper, we do not aim to present a better DR algorithm and examine its impact. Rather, we implement an already established DR algorithm from Mengelkamp et al. (2018a) and then represent the impact of changing level of DR on trading behaviour and reinforcement learning of the participants in the LEM since there is negligible literature which studies the impact of DR strategies on reinforcement learning of trading strategies. The target of the paper is to study three aspects. First to study the impact of changing level of DR on learning and economic benefit of the participants of the LEM. Second to determine the variation of parameters in the modified Erev-Roth algorithm to determine different trading techniques for the participants. Third to analyze the impact of increasing the share of RES in power generation on the LEM. The paper tries to bridge the gap between the three aspects of peer-to-peer trading, DR and reinforcement learning and its impact on each other to establish a LEM which provides not only economic benefits and partial self-sufficiency to its participants but also provide grid flexibility to the DSO and also curtails the capital expenditure of deploying electricity generators to meet the growing demand of the electricity.

Methodology and model

The model we have used for the sensitivity analysis of DR in LEM is adapted from the model used in Mengelkamp et al. (2018a). We have repeated the description of the model here so that the readers does not have to switch papers to understand the working of the model.

Agent definition

A community of 100 residential households is represented in an agent-based model that incorporate a LEM functioning on peer-to-peer trading through a short term 15 min time-slot based merit order market mechanism. There are different kinds of households as represented by the agents i.e. prosumers and consumers. Prosumer agents are those agents who have their own electricity generation unit (e.g. PV or mCHP). Consumer agents are those who do not have their own generation units and thus depend on trading or the grid for their electricity supply. Apart from these agents, there is also the market maker which is represented as market agent in the model that receives the bids and offers from the household agents, matches the bids and offers according to the merit order mechanism and then sends back the information about the successful bids and offers to the corresponding agents.

Model description

The household agents send their bids and offers based on the pricing strategy for the next 15 min to the market agent. The market agent sorts the bids and offers in decreasing and increasing order respectively to establish the demand and supply curves. These curves are used for matching according to the merit order market mechanism. The intersection of the demand and supply curve determines the market closing price (MCP) for that particular time-slot and all the trades accepted buy and sell their energy at this uniform price for that time-slot. The information about the successful trades is sent back to the respective agents and the pricing strategy is updated accordingly for the next 15 min time-slot.

   Each household agent executes its pricing and DR strategy on an individual basis. The pricing strategy of the agents is based on the modified Erev-Roth algorithm (Erev and Rapoport 1998; Nicolaisen et al. 2001) and explained in detail in the "Pricing strategy" section and the DR is explained in the "Demand shifting in DR" section. In the model, the market clearance done by a trusted third party is not an agency, an individual or a company. Rather, the innovation in the field of IoT makes this job easier because in our model the processes explained can easily be taken care by a device which can receive the bids and offers from different households, sort them out accordingly and match them as per the merit order market model. In this regard, blockchain technology can act as an added layer of security and trust for recording the transactions as explained in an actual LEM established in Landau, Germany by Mengelkamp et al. (2018d). In this paper, we have not investigated the application of blockchain in LEM.

Pricing strategy

The application of reinforcement learning is described through various literature in "Reinforcement learning in LEM" section. In this section, we implemented the modified Erev-Roth algorithm to develop the pricing strategy of the model. The pricing strategy is aimed at increasing the individual economic benefit of the agents in the LEM. The minimum and maximum bid and ask prices are based on existing price components in the German retail electricity market, which represent a natural alternative to trading on a LEM.

   A set of strategies $S = \{s_1, s_2, ...s_m\}$ for each individual agent i is set up which correspond to the discreet bids (or offers) an agent will execute in the LEM. Initially, the agents have no prior knowledge about the behaviour on the LEM except for the upper (c^G) and the

lower (c^F) limits of the trading window. S correspond to all the bids or offers between (c^G) and (c^F) with increment at discrete c€ level with one decimal point. Initially, at t_0, the propensity $q_{is}(t)$ of all the strategies s for an agent i are equal and set by Eq. (1).

$$q_{is}(t_0) = sca(t_0) \frac{avg(\Pi_i(t_0))}{|S|}. \tag{1}$$

$avg(\Pi_i(t_0))$ is the profit earned by an individual agent i for the time-slot t_0 and $sca(t_0)$ is the scaling parameter. After the round of trading at time-slot t among various agents, the agents update their propensity for the next time-slot $(t + 1)$ to bid for the time-slot $(t + 2)$ through Eq. (2).

$$q_{is}(t + 1) = (1 - rec)q_{is}(t) + MUF(i, s, s', t). \tag{2}$$

The recency effect of past events is determined by *rec* parameter (Erev and Rapoport 1998) and the modified update function (*MUF*) is given by Nicolaisen et al. (2001). It is based on the chosen strategy s' at time-slot t which is given by Eq. (3).

$$MUF(i, s, s', t) = \begin{cases} \Pi_i(t)(1 - exp), & s = s' \\ q_{is}(t) \frac{exp}{|S|-1}, & s \neq s' \end{cases}. \tag{3}$$

The *exp* parameter reduces the propensities of the not chosen strategies and also actuate the weightage of the current strategy on the profit (Erev and Rapoport 1998). The probability for a certain strategy s is then determined by Eq. (4).

$$p_{is}(t + 1) = \frac{q_{is}(t + 1)}{\sum_{s=s_1}^{|S|} q_{is}(t + 1)} \tag{4}$$

Initially at time t_0, the probabilities for all the strategies are equal and determined by $p_{is}(t_0) = 1/|S|$. After the first market clearance, when all the individual agents have chosen their strategies randomly, the modified Erev-Roth algorithm comes into play and determine the probabilities of the future bids and offers and gets updated according to the success or failure of the chosen strategies.

Demand shifting in DR
The demand shifting is based on a strategy as presented in Mengelkamp et al. (2018a). The demand profile of the individual agents $I = \{1, 2, \dots, N\}$ is forecasted perfectly for the next 24 h at 15 min interval as $D_i = \{d_{i,t_0}, d_{i,t_1}, \dots, d_{i,t_{95}}\}$. The maximum peak of the forecasted demand is determined by the maximization function $D_i(d_{i,max})$ as Eq. (5).

$$D_i(d_{i,max}) = max(D_i(d_{i,t})), t = t_0, t_1, \dots, t_{95}. \tag{5}$$

Then a parameter *SDR* is determined based on a perfect foresight as to how much proportion of the maximum peak in a day can be shifted to a new time interval. The assumption of perfect forecast to develop an LEM model based on reinforcement learning is taken from the support of the paper (Mengelkamp et al. 2018a). The *SDR* is defined in the range of [0,1], where 0 represents the whole peak should be shifted and 1 represents no DR at all. The load shifting is applied to all those points of the load curves which satisfies the Eq. (6).

$$D_i(d_{i,t}) > D_i(d_{i,max}) * SDR, t = t_0, t_1, \ldots, t_{95}. \tag{6}$$

The expression $D_i^{shift}(d_{i,t(j)}), j = 1, 2, ..n$, represents the intervals of the load curve which satisfies the Eq. 6, where n is the number of peaks above the *SDR* limit, and the values of $D_i^{shift}(d_{i,t(j)})$ are given by Eq. (7).

$$D_i^{shift}(d_{i,t(j)}) = D_i(d_{i,t(j)}) - D_i(d_{i,max}) * SDR, \tag{7}$$

where $t(j) = t(j_1), t(j_2), \ldots, t(j_n)$. The load intervals of a particular agent i which are above the *SDR* is denoted by index j.

The minimum demand of the forecasted demand profile $D_i(d_{i,t})$ for 24 hours interval is denoted by $D_i(d_{i,min})$ and the time at which the minimum demand for a particular agent takes place is denoted as $t_{i,min}$. The $D_i(d_{i,min})$ is calculated through the Eq. (8).

$$D_i(d_{i,min}) = min(D_{i,t}), t = t_0, t_1, \ldots, t_{95}. \tag{8}$$

The demand $D_i^{shift}(d_{i,t(j)})$ that is to be shifted as determined by Eq. (7) is then moved to the time interval when demand is minimum $D_i(d_{i,min})$ at the time interval $t_{i,min}$ and added to $D_i(d_{i,min})$. This step generates a new demand profile $D_{i,t}^{new}(d_{t_0}, d_{t_1}, \ldots, d_{t_{95}})$. Once this step is iterated for 96 times the final demand profile is set and denoted as D_i^{final} with the goal of reducing the peak demand of the individual agents of the LEM. This modified demand profile D_i^{final} is sent back to the individual agents for trading in the LEM.

Key performance indicators (KPIs)

We have determined certain KPIs to analyse the technical and economic aspects of trading and DR facilitated with RL on the LEM and the physical grid on which the LEM is embedded. The KPIs facilitate the analysis of our chosen regulatory scenarios and the sensitivity analysis of the effect of change in the degree of DR and installed generation capacity on the LEM.

The KPIs we apply are:

1.   Degree of Local Sufficiency (DLS)
2.   Market Closing Price (MCP)
3.   Residual Peak Demand (RPD)

Mengelkamp et al. (2018a), de Oliveira e Silva and Hendrick (2017), and Long et al. (2018) have used the DLS as one of the KPIs to evaluate the efficiency of the microgrid model. Mengelkamp et al. (2018a), Zhou et al. (2020) and Chen and Bu (2019) have used the MCP to determine the economic benefits and setup the constraints for their learning algorithm of their model. Mengelkamp et al. (2018a) and Marzband et al. (2013) have utilised RPD to evaluate the optimum flexibility that can be offered by a LEM to a transmission grid. Since, we wanted to explore the paradigm of efficiency, economic benefits and flexibility offered by an LEM, we chose the above mentioned three KPIs for our study. Apart from these KPIs, there are several other KPIs mentioned in existing literature, however, it is out of the scope of our study.

The DLS is defined as the ratio of the total consumption of generated electricity which includes the energy self-consumed $sc_{i,t}$ or traded $et_{i,t}$ among the agents $i \in I$ in the LEM

to the total aggregated original demand $\sum_{i=1,t=t_0}^{N,T} d_{i,t}^{original}$ without any DR of the LEM. The DLS is given by Eq. (9).

$$DLS = \sum_{i=1}^{N} \sum_{t=t_0}^{T} \frac{sc_{i,t} + et_{i,t}}{d_{i,t}^{original}}. \tag{9}$$

The MCP helps to analyse the net profit the LEM gains from peer-to-peer trading, in comparison to buying from the grid and selling to the grid. The MCP is defined as the weighted average of the market clearing price that happens every 15 min over the year and given by the Eq. (10).

$$MCP = \frac{\sum_{i=1,t=t_0}^{N,T} MCP_t et_{i,t}}{\sum_{i=1,t=t_0}^{N,T} et_{i,t}}. \tag{10}$$

The RPD is defined as the aggregated residual annual peak demand of all the household agents after self consumption, trading energy and DR in the LEM. It determines the maximum peak of demand for the LEM that has to be supplied by the grid. The RPD is denoted by $D_{i,t}$ and given by the Eq. (11).

$$RPD = \max \begin{cases} \sum_{i=1}^{N} D_{i,t_0} - (sc_{i,t_0} + et_{i,t_0}); \\ \sum_{i=1}^{N} D_{i,t_1} - (sc_{i,t_1} + et_{i,t_1}); \\ \qquad \cdots \\ \sum_{i=1}^{N} D_{i,T} - (sc_{i,T} + et_{i,T}). \end{cases} \tag{11}$$

Set of scenarios

We distinguish our set of scenarios firstly concerning the regulatory context of Germany and secondly by the degree of interaction among the agents. We have defined three types of regulatory scenarios:

1. Public Network: virtual community on the (national) grid level
2. Microgrid: real community on a local perimeter
3. Favorable Regulation: idealised scenario

The regulatory scenarios determine the lower price limit (c^F) of trading electricity in the LEM. The virtual community on the national grid level takes into account the full regulatory cost of peer-to-peer trading, including (renewables) surcharges, taxes, and network and concession fees. The Microgrid scenario is based on based on the regulatory concept of a customer installation (Kundenanlage) in Germany, where a limited number of peers can trade electricity among each other inside a local perimeter without paying grid fees and electricity taxes. The Favorable Regulation is an idealised regulation where apart from grid and electricity tax relaxation, the community is also exempted from the renewable surcharges. The upper limit (c^G) in this trading window is based on a reference tariff for the 2018 retail grid electricity price. The lower limit (c^F) of the trading window is based on the feed-in tariff of PV and mCHP. Corresponding taxes and regulatory surcharges are taken into account while calculating the limits of the trading window as adapted from Mengelkamp et al. (2018a). The upper and lower limits of the trading window for various scenarios are given in Table 1.

In the Public Network scenario, trading is not economic since the lower limit (c^F) of the trading window is higher than the upper limit (c^G) of the trading window as can be seen from Table 1. Trading is economically not beneficial because there are various surcharges

Table 1 Upper and lower limit of the trading window (2018 prices in c€/kWh)

Regulatory scenarios	(c^G)	(c^F)	
		PV	CHP
Public Network	29.86	36.74	35.81
Microgrid	29.86	25.01	24.09
Favorable Regulation	29.86	16.83	15.90

and taxes that are levied on selling electricity through the national grid. The detailed price description of various costs in the above mentioned regulatory scenarios can be found in Mengelkamp et al. (2018a).

The second set of scenarios is based on the degree of technical interaction among the different agents in the LEM. We have defined four types of scenarios which are :

1.   Base Case
2.   Trading
3.   Trading & DR
4.   Upper Bound (i.e Trading + DR + UL)

In the base case, there is no application of trading or DR among the household agents. The trading scenario depicts the case when there is peer-to-peer trading among the agents facilitated by RL.The trading & DR case incorporates DR of individual agents on top of peer-to-peer trading. The upper bound case is a case of peer-to-peer trading supported by DR but the bids of electricity are set to grid price i.e. all the electricity in the LEM are asked at a price equal to the price the agents would have to pay while buying from the grid i.e Upper limit of the trading window (c^G). The interaction of agents is described in table as given in Mengelkamp et al. (2018a).

Simulation setup

The set up of the market from "Methodology and model" section is implemented into an agent-based model using the Anylogic software. The Main class of the model initiates all other agents with prosumer and consumer population of agents along with the demand and generation curves for each household agents and simulation time-slot is set at 15 min intervals for 1 year. The prosumer or consumer population of agents execute the pricing strategy and the DR strategy, and the constructed bids and offers are sent to the market clearing agent for clearance. Once the trades are matched through merit order model, the information about successful trades are sent to the household population of agents. A detailed setup of the simulation can be found in Mengelkamp et al. (2018a). The implementation of the regulatory scenarios are actualized using the upper(c^G) and lower limit (c^F) of the trading window as given in Table 1. The trading scenarios are as follows:

1.   **Base**: The pricing and the DR strategy is switched OFF in this case.
2.   **Trading**: The pricing strategy is switched ON but the DR strategy is switched OFF in this case.
3.   **Trading+DR**: Both the pricing and the DR strategy is switched ON in this case.
4.   **Trading+DR+UL**: Both the pricing and the DR strategy is switched ON here. In addition, the excess electricity that is generated and sold in the LEM is bought at grid price (c^G) to enforce the selling of all the local electricity generated in the LEM.

The simulations are run for 1 year and an evaluation function reports the KPIs, that are calculated to analyse the performance of the LEM.

Simulation results

Data origin

The PV data is obtained from a PV installation in the Southern part of Germany which is recorded at 15 min time-slots for 1 year. The generation curves for the prosumer households is then obtained from this curve using a 20% uniform distributed randomization function. The mCHP generation data is obtained from averaging multi-year data of 9 mCHP installations (1 in Southern Germany, 1 in Alsace (France) and 7 in Fortainbleau (France)) of 0,7-1 kWp installed electric power. The consumption profiles of households are obtained from Unna (2002) and the curves are uniformly distributed as in PV generation data to fit 1–5 person households.

Test runs

A set of 10 test runs in 15 min time-slots for 1 year is run for the model for every scenario. The pricing strategy and the DR strategy is switched ON or OFF based on every case and the pricing strategy is initialized with parameter values *sca* = 1,0, *rec* = 0,02, *exp* = 0,99 from Nicolaisen et al. (2001). The test runs are conducted on a standard laptop with Processor Intel(R) Core(TM) i5-7300U CPU @ 2.60GHz, 2701 Mhz, 2 Core(s), 4 Logical Processor(s) along with 16.00 GB RAM. The simulation is executed in the Anylogic University Researcher Edition 8.5.1 software. One simulation run takes on an average 10 min to complete 1 year in 35040 time-slots.

Sensitivity analysis

A sensitivity analysis of all the 12 scenarios is done on two metrics. In the first metric, the PV peak power installation is increased from 5kWp to 25kWp in 5kWp intervals (i.e 5kWp, 10kWp, 15kWp, 20kWp, and 25kWp). The second metric that is chosen is the DR % which has a range from 0% to 50% in 10% intervals (i.e. 0%, 10%, 20%, 30%, 40%, and 50%) which corresponds to *SDR* of value (100%, 90%, 80%, 70%, 60%, and 50%) respectively. This creates a matrix of 60 cases for each combination of scenarios which is used for sensitivity analysis of the performance of the LEM that is analysed through the KPIs.

Evaluation

Evaluation of the modified Erev-Roth algorithm

In order to study the impact of the parameters of the modified Erev-Roth algorithm, we made some tests and focused on the evolution of strategy in time, and on the gain generated from the energy trading compared to buying energy at the grid price. The evaluation was done for the scenario of favorable regulation scenario with a fixed DR of 30%. In order, to evaluate the algorithm, we defined certain performance indicators:

1.   Average profit: the accumulated profit for a certain bid price
2.   Strategy: the bid price associated to average profit
3.   Gain from trading: the accumulated money saved from trading

We plotted these values for different values of *rec* (always for the same Household), for 1000 h. The value of exp is given in the paper from Nicolaisen et al. (2001), and is set

to 0,99. For the rest of the paper, the following nomenclature is followed for the time t_c required by the algorithm to converge to a certain strategy:

1. Fast convergence: $t_c < 150$ h
2. Moderate convergence: 150 h $< t_c < 500$ hours
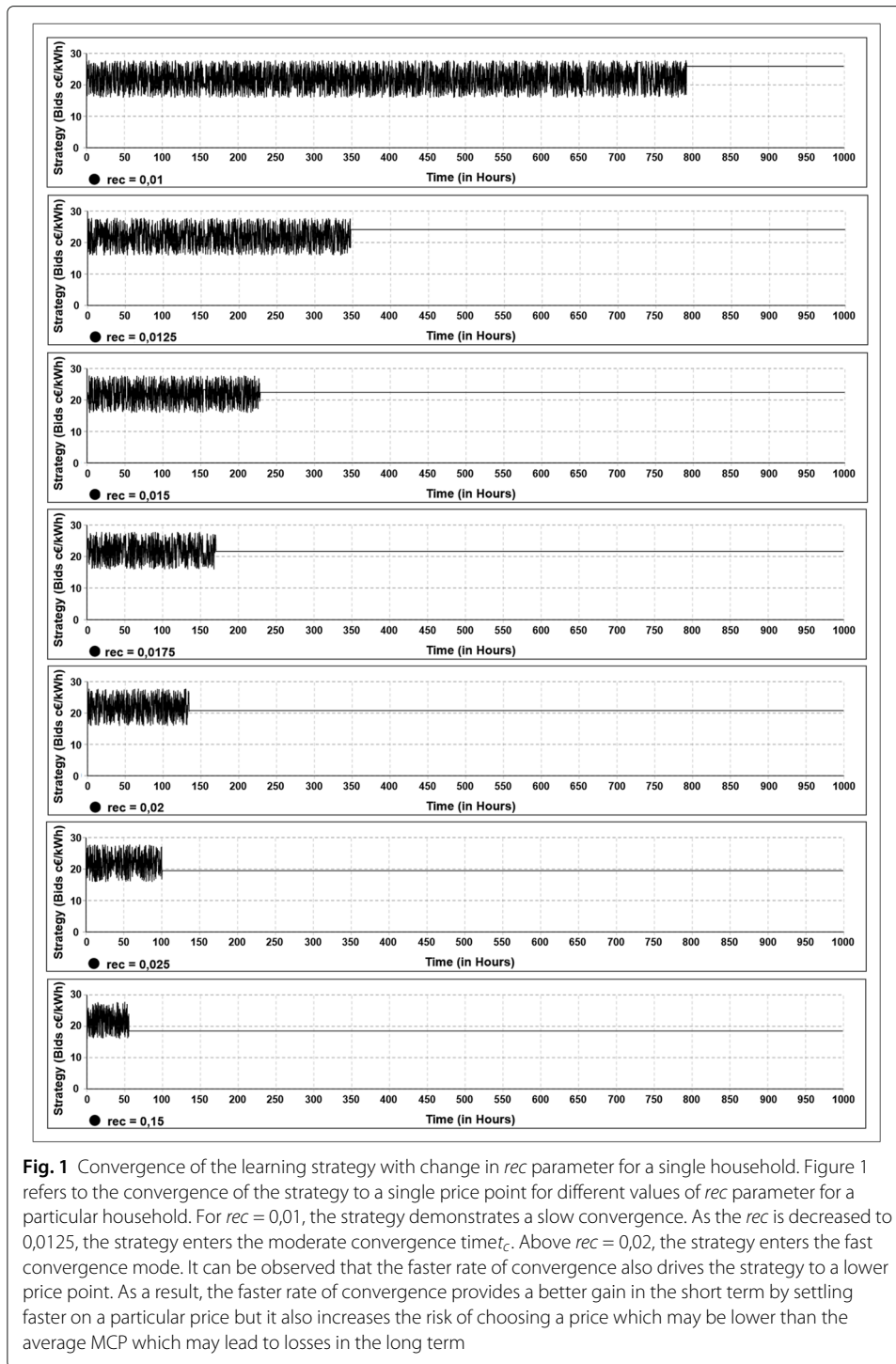3. Slow convergence: $t_c > 500$ h

The rate of convergence for various values of *rec* is demonstrated in Fig. 1. For *rec* = 0,01, the strategy converges to a constant value between 500h and 1000h(slow convergence). It can be seen that it converges to a "safe" value, because this bid price will be usually higher than the MCP, so many bids will be accepted. As we increase *rec,* the time of convergence slowly reduces along with the bid price. For *rec* = 0,0125, the strategies show moderate convergence. Also, it converges to a lower value which is a bit riskier compared to the strategy converged at *rec* = 0,01, but it is still sufficiently above the MCP. As we keep on increasing the *rec* parameter at *rec* = 0,02, the strategy shows fast convergence. In this case, the price at which the strategy converges is near to the annual average MCP which poses a risk of choosing a wrong strategy if the MCP goes above the value of the converged strategy. Above, *rec* = 0,02, the strategies tends to keep on converging at a much faster pace. However, the converged strategy falls substantially below the annual average MCP which may cause substantial risk of lower gains. To understand the development of gains from converging strategies for various values of *rec,* the gain from trading was plotted against time and demonstrated in Fig. 2.

The lower values of *rec* parameter correspond to long term strategies. The value of *rec* = 0.01 shows lower gains in the beginning but increases at a faster rate than other strategies. As we increase the *rec* parameter to 0,015 and 0,0175, it can be observed that the gains are better than other strategies for mid-term. This approach seems to be very good to mix safety in long term and good income in short term. The *rec* parameter at 0,0175 seems to be efficient and can satisfy an individual who is ready to take risks in order to have a big and fast income. For $rec = 0,02$, shows similar gains as that of *rec* at 0,015 in short term but tends to fall bellow all other strategies in long term. From here, it can be concluded that the rate of increase of this curve can be linked to the value of *rec,* and the rate is higher for lower values of *rec,* but also riskier than the ones with higher values of *rec.*

Evaluation of the results of the model
The sensitivity analysis is performed on all the combinations of regulatory scenarios and the scenarios based on interaction of agents. We intend to evaluate the impact of three different regulatory scenarios on each KPI separately (i.e. DLS, MCP and RPD). The lowest values of KPIs for Fig. 3 are marked in red colour and the highest values are marked in blue colour and vice-versa for Fig. 5. The intermediate values are marked according to their closeness to the two extreme values.

Figure 3 demonstrates the sensitivity analysis of the degree of local sufficiency of the LEM. The application of DR has positive impacts on the DLS on all regulatory scenarios. The public network scenario does not provide any window for trading. The increase in DR% however increases the DLS by 19–29% by increasing the PV installation from 5–25kWp. The microgrid and the favourable regulation provide window for trading and the performance of DLS is further increased through increase in DR%. However, in the relative comparison of the Microgrid and the Favorable regulation scenarios where

Fig. 1 Convergence of the learning strategy with change in *rec* parameter for a single household. Figure 1 refers to the convergence of the strategy to a single price point for different values of *rec* parameter for a particular household. For *rec* = 0,01, the strategy demonstrates a slow convergence. As the *rec* is decreased to 0,0125, the strategy enters the moderate convergence time t_c. Above *rec* = 0,02, the strategy enters the fast convergence mode. It can be observed that the faster rate of convergence also drives the strategy to a lower price point. As a result, the faster rate of convergence provides a better gain in the short term by settling faster on a particular price but it also increases the risk of choosing a price which may be lower than the average MCP which may lead to losses in the long term

trading and DR is implemented, it can be observed that the DLS increases by 20% in Microgrid scenario and 15% in Favorable regulation scenario for 5kWp as we increase DR% from 0% to 50%. The upper bound scenario demonstrates maximum level of DLS for all the regulatory scenarios since these particular scenarios enforce maximum trading of electricity in the LEM. In the scenarios involving trading with DR, the agents do

Fig. 2 Gain from trading with change in *rec* parameter for a single household. Figure 2 refers to the gains obtained from trading for different values of the *rec* parameter. For *rec* = 0,01, the gain from trading starts a bit lower than most of the other sets but the shape of the curve suggests that it increases at a faster rate with time. For *rec* = 0,0125, we can see that it starts a bit higher than with the previous set of parameters, but gets lower between 500h and 1000h. This is a short-term strategy, but we can see by comparing to the other curves that it does not seem to be efficient, as it is almost always lower than most of the other curves. For *rec* = 0,015, the gain starts at a high value. Moreover, it increases at a faster rate, and we can see that after 1500h it is higher than any other curve. The gain obtained from *rec*=0,0175 seems to be highest for short term. However, as time passes, it tends to slowly fall below the other curves. For *rec* = 0, 02, the gain in the beginning is roughly the same as *rec* = 0.015. However, for long term, it can be seen that the curve with this set increases much slower than the other

not bid for all the energy, rather they bid intelligently as per the pricing strategy of individual agents. The extension of trading window through regulatory scenarios does not necessarily increase the DLS as observed in the similar cases where the purchase of local generation is not enforced for corresponding cases of Microgrid and Favorable regulation. The increase of PV installation and increase in percentage of *SDR* both have positive impact on the DLS. In addition, as the regulatory barriers decrease which in turn broadens the trading window, leads to an increase in the DLS. However, the maximum % of DLS that can be achieved in a case similar to our LEM is about 81%.

de Oliveira e Silva and Hendrick (2017) provides the demonstration of self-sufficiency of 25 Belgian households using Lithium-ion batteries. A self-sufficiency of 30% was achieved using only PV installation of 5kWp. Above that, storage was used to achieve a self-sufficiency of 80%. Long et al. (2018) described a model of a microgrid with 100 households out of which 40% of the households had their own PV generator along with battery storage. A self-sufficiency of 33,7% was achieved with peer-to-peer trading without any battery storage which increased up to 47,4% with the use of 16kWh of storage for those prosumer households. In comparison, we achieved DLS of 22,1% with PV installation. We replaced the battery with DR and were able to reach up to 36,6%, an increase of 14,5% with 30% DR. A maximum of 48,6%, an increase of 26,5% was achieved with the implementation of 50% DR.
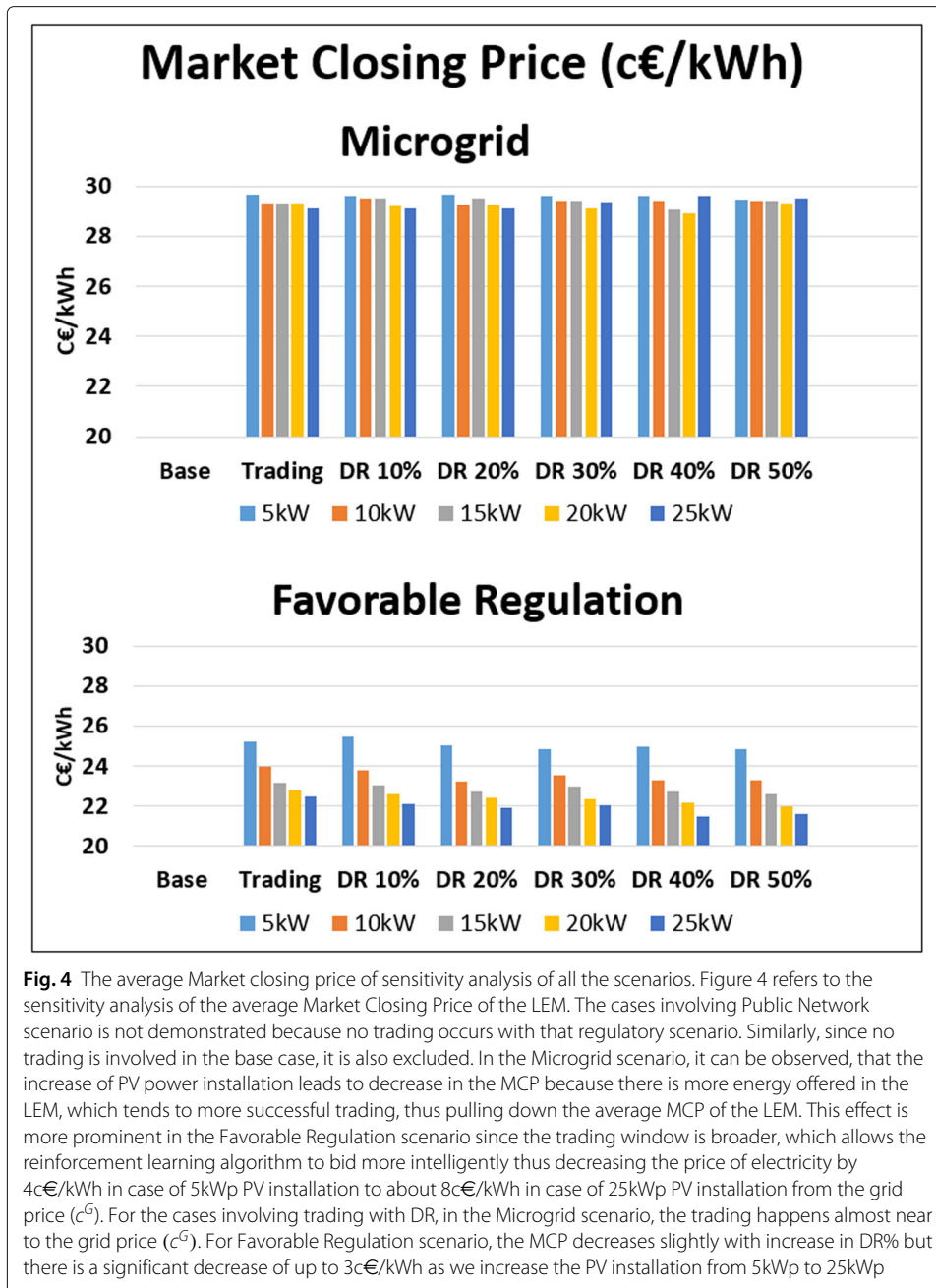
Figure 4 displays the sensitivity analysis of the annual average MCP for all the combination of scenarios. For the cases involving trading along with DR with or without enforcement of the prices of local generation of PV at grid price (c^G) in the Microgrid scenario, the trading happens almost near to the grid price (c^G) because of a small trading window and all the bids are forced to be equal to grid prices. Increasing the DR% to

DLS(%)		Public Network					Microgrid					Favorable Regulation				
PV(kW)→ DR(%) ↓		5	10	15	20	25	5	10	15	20	25	5	10	15	20	25
Base	0%	22,1	25,5	26,9	27,7	28,2	22,1	25,5	26,9	27,7	28,2	22,1	25,5	26,9	27,7	28,2
Trad.	0%	22,1	25,5	26,9	27,7	28,2	22,3	26,3	28,0	29,0	30,0	31,4	40,9	47,0	51,7	55,1
	10%	31,9	38,1	40,8	42,5	43,6	31,7	38,0	40,9	43,2	45,1	37,8	49,2	57,5	62,4	66,8
Trad	20%	34,5	41,3	44,5	46,3	47,5	34,5	41,7	44,8	47,1	49,2	40,4	52,3	60,8	65,5	70,3
+	30%	36,6	43,9	47,4	49,4	50,9	36,6	44,1	47,6	50,6	51,5	42,3	54,6	62,1	67,6	72,6
DR	40%	38,9	46,8	50,4	52,6	54,2	38,9	46,8	51,3	53,8	54,2	44,1	57,2	64,5	71,5	76,0
	50%	41,2	49,4	53,4	55,7	57,4	41,4	49,5	53,5	56,2	57,5	46,6	60,1	68,0	73,7	79,2
Trad.	10%	31,9	38,1	40,8	42,5	43,6	41,0	52,4	59,4	65,5	70,2	40,9	52,2	59,3	65,4	70,6
+	20%	34,5	41,3	44,5	46,3	47,5	42,8	54,6	62,6	68,9	73,7	42,7	54,4	62,4	68,9	73,7
DR	30%	36,6	43,9	47,4	49,4	50,9	44,6	56,9	65,2	71,4	76,3	44,4	56,7	65,3	71,7	76,9
+	40%	38,9	46,8	50,4	52,6	54,2	46,6	59,3	68,1	74,5	79,1	46,4	59,1	68,1	74,5	79,0
UL	50%	41,2	49,4	53,4	55,7	57,4	48,6	61,9	71,1	77,4	81,3	48,4	61,6	70,8	77,6	81,8

Fig. 3 The Degree of Local Sufficiency of sensitivity analysis of all the scenarios. Figure 3 refers to the sensitivity analysis of the degree of Local Sufficiency for all the combination of technical and regulatory scenarios. The undesirable sensitivity i.e low DLS is denoted in red color, and as the DLS increases through the heat map, the desirable DLS i.e a high DLS in demonstrated in green color with intermediate values denoted with a color transition from red to green with increasing DLS. It can be observed that the DLS increases with both increasing the local PV production and also by increasing the degree of DR. In the Public Network scenarios without DR, the DLS increases by 6,1% from 22.1% to 28,2% as we increase the PV installation from 5kWp to 25kWp because more energy generation leads to more self-consumption. The implementation of trading in Favorable Regulation scenario provides a bigger window for trading which leads to increase in consumption of electricity generated locally in the LEM, which in turn, increases the DLS by 9,3% for 5kWp installation to 26,9% for 25kWp installation in comparison to the corresponding base case scenarios. In the Microgrid scenario, the DLS increases by 19,1% for 5kWp and 27,4% for 25kWp installation and in Favorable regulation scenario, the DLS increases by 15,2% for 5kWp and 24,1% for 25kWp as we increase DR% from 0%–50%. For the cases where purchase of local generation is enforced, the Public Network scenario shows the increment of DLS without trading to 41,2% for 5kWp PV installation to 57,4% for 25kWp installation. The effect of trading can be visualized in case of the Microgrid scenario in combination with Upper bound scenario where the DLS increases by 13% for 5kWp to 25,9% for 25kWp installation. However, a further relation of regulations does not increase the DLS as can be observed in Favorable regulation cases in comparison with the corresponding Microgrid cases

50% also does not have a significant impact on the MCP. For Favorable Regulation scenarios, however, there is a significant decrease of the MCP up to 3c€/kWh as we increase the PV installation from 5kWp to 25kWp due to the presence of more offers of electricity, which enables the intelligent agent strategy to lower the price of electricity in the LEM. For the cases involving enforcement of prices of local generation of PV at c^G, the price settles around 27c€/kWh, thus decreasing the average price of electricity by 2c€/kWh and it is not much affected by increase in PV power installation or increase in DR, which showcases the fact that if the bids are fixed to the grid price to ensure maximum consumption of locally generated electricity, the MCP decreases by a small margin but it is not substantially affected by DR.

When comparing our results with existing literature, it is observed that Zhou et al. (2020) explored the paradigm of user dominated DR and peer-to-peer trading on a local energy market of 50 households. Here, a PV installation of 3,2 kWp was used for the simulation. With a penetration of 50% PV, which is similar to our case, an annual saving on the cost of electricity provision of 17,7% was achieved with only peer-to-peer trading. The increase in savings of the consumers through DR was not reported by Zhou et al. (2020).

Fig. 4 The average Market closing price of sensitivity analysis of all the scenarios. Figure 4 refers to the sensitivity analysis of the average Market Closing Price of the LEM. The cases involving Public Network scenario is not demonstrated because no trading occurs with that regulatory scenario. Similarly, since no trading is involved in the base case, it is also excluded. In the Microgrid scenario, it can be observed, that the increase of PV power installation leads to decrease in the MCP because there is more energy offered in the LEM, which tends to more successful trading, thus pulling down the average MCP of the LEM. This effect is more prominent in the Favorable Regulation scenario since the trading window is broader, which allows the reinforcement learning algorithm to bid more intelligently thus decreasing the price of electricity by 4c€/kWh in case of 5kWp PV installation to about 8c€/kWh in case of 25kWp PV installation from the grid price (c^G). For the cases involving trading with DR, in the Microgrid scenario, the trading happens almost near to the grid price (c^G). For Favorable Regulation scenario, the MCP decreases slightly with increase in DR% but there is a significant decrease of up to 3c€/kWh as we increase the PV installation from 5kWp to 25kWp

Long et al. (2018) have reported similar findings for a microgrid with 100 households with 40% households equipped with PV panels and battery storage. An annual decrease of 30% cost of electricity was reported for the community through peer-to-peer trading. Chen and Bu (2019) has explored the self-learning prosumer behaviour of developing intelligent agent strategies through deep reinforcement learning method in a LEM of 200 households. The average annual revenue saved through this method for the LEM with only trading was reported as 33% saving with trading in LEM and 54% with trading and storage. In comparison to our case, the average annual MCP in Microgrid scenario was reduced by 0,7c€/kWh (2,3%) and by 4c€/kWh (13,4%) in the Favorable Regulation scenario for

the consumers for a 5kWp PV installation on 45% of the households. Our intelligent agent strategies managed to achieve annual average saving of 7,56c€/kWh (25,3%) with only 10% DR. As for prosumers, they made a profit with average annual MCP of 25,4c€/kWh with trading instead of putting it in the grid and achieving a feed-in tariff 16,83c€/kWh i.e. 51% annual increase in the revenue which is comparable to that of Chen and Bu (2019). However, it must be noted that the results from Chen and Bu (2019) corresponds to the regulations of United States of America but in our case, it corresponds to Germany. This may have a difference in the trading window and may impact the results as well.

The annual RPD is demonstrated in Fig. 5. The effect of only DR without trading can be observed for scenarios involving trading with DR in the Public Network scenario. A slight increase of DR% by only 10% can decrease the RPD by 22–25% because demand shifting can move the load to definite time-slots, because the DR strategy utilized for this model is price-based and shift of load from evening to morning can have significant impact on decreasing the RPD of the LEM. However, there are sudden peak surges for cases with DR more than 30% because of excessive demand shifting leads to local maxima in the load curves. This problem of sudden peaks is mitigated, when we move from Public Network scenarios to Microgrid or Favorable Regulation scenarios which involves trading with DR. Another interesting outlook is that the increase of trading window has negligible impact on the RPD which can be observed by comparing the corresponding cases of trading with DR in the Microgrid and Favorable Regulation scenarios.

RPD(kW)		Public Network					Microgrid					Favorable Regulation				
PV(kW)→ DR(%) ↓		5	10	15	20	25	5	10	15	20	25	5	10	15	20	25
Base	0%	16,2	16,0	15,9	15,8	15,8	16,2	16,0	15,9	15,8	15,8	16,2	16,0	15,9	15,8	15,8
Trad.	0%	16,2	16,0	15,9	15,8	15,8	16,2	16,0	15,9	15,8	15,8	16,2	15,9	15,7	15,6	15,5
Trad + DR	10%	12,1	11,7	11,9	12,3	12,2	9,3	9,3	9,3	9,6	10,0	9,5	8,8	8,8	8,7	8,8
	20%	11,7	13,7	14,1	15,4	15,7	8,9	8,4	8,5	8,4	9,0	10,1	7,5	7,7	7,6	9,5
	30%	9,4	8,4	9,0	9,5	9,4	8,4	8,1	8,1	8,5	9,4	7,8	7,6	7,1	7,3	9,3
	40%	24,2	23,9	23,8	24,2	24,7	7,9	7,8	8,4	9,0	8,5	7,4	7,7	7,2	9,3	8,9
	50%	11,2	11,4	11,4	10,8	11,9	7,9	7,6	7,7	10,1	9,5	7,0	7,9	8,1	10,0	10,0
Trad. + DR + UL	10%	12,1	11,7	11,9	12,3	12,2	9,2	8,8	8,5	8,9	8,5	9,5	8,8	8,5	8,3	8,0
	20%	11,7	13,7	14,1	15,4	15,7	8,4	7,3	7,0	7,2	9,0	10,1	7,4	7,6	7,1	9,5
	30%	9,4	8,4	9,0	9,5	9,4	7,0	7,0	6,8	7,3	8,3	7,7	7,6	6,5	7,0	9,2
	40%	24,2	23,9	23,8	24,2	24,7	6,7	7,2	7,4	7,9	7,5	6,3	7,7	7,0	9,2	8,7
	50%	11,2	11,4	11,4	10,8	11,9	6,4	6,3	7,2	9,4	8,8	6,6	7,5	8,0	9,9	9,9

Fig. 5 The sensitivity analysis of residual Annual peak demand of all the scenarios. Figure 5 refers to the sensitivity analysis of Annual Residual Peak Demand of the LEM. The undesirable sensitivity i.e high RPD is denoted in red color, and as the RPD decreases through the heat map, the desirable RPD i.e a low RPD is demonstrated in green color with intermediate values denoted with a color transition from red to green with decreasing RPD. In the base cases of Public Network scenario involving no DR, the increase of PV installation from 5kWp to 25kWp decreases the RPD by 2,5% which demonstrates the fact that the most surges in electricity demand occurs near or after sunset. A slight increase of DR% by only 10% can decrease the RPD by 22–25% in the trading with DR scenario in the Public Network scenario. However, if we keep increasing the *SDR* there is decrease in the RPD up to a certain limit after which the RPD surges significantly if DR% is more than 30% as too much peak shading and movement of load curves causes formation of local maxima which leads to sudden peaks in the RPD. In the Microgrid scenario combined with trading and DR, the increase of PV power installation does not impact the RPD much for lower percentage of DR. However, as we keep increasing the percentage of DR, the RPD has significant drop of up to 35–42%. In the Favorable Regulation scenario combined with trading and DR, there is negligible decrease in the RPD when compared to the corresponding cases in the Microgrid scenario

Discussion

The analysis of changing the *rec* parameter changes the behaviour of the model, and more specifically the time of convergence of the bidding strategy. The time of convergence also influences the evolution of the gain from trading. With the set of parameters that makes a fast convergence strategy (t_c<150h), gain is strong at the beginning. However, after thousand hours, moderate converging strategies (150h<t_c<500h) seems to be more efficient, as the gain increases faster. Slow converging strategies (t_c>500h) seem to be interesting on the very long term because the gain at the beginning is lower in comparison to other strategies. It was also noticed that the strategy chosen with quick converging parameters is often riskier than the parameter corresponding to slower convergence. This induces a higher gain in the short term but can also be a loss making strategy if many bids are rejected because it converged considerably below the MCP.

The sensitivity analysis of the combination of scenarios demonstrates how the DLS, MCP and RPD changes with change in PV power generation and change in percentage of peak shading in DR. The DLS can reach above 80% with increase in PV power installation. A similar development of DLS can be observed with increasing the DR% and a substantial gain of around 40–50% can be achieved even for PV installation of 5kWp. The range of trading window has a significant impact on the average MCP of the LEM.

Our analysis shows that the introduction of LEM, if set up in a convenient way for the participating agents, could prove to be a practical solution for maximization of local value generation in an increasingly decentralised energy system based on renewable energy sources. The microgrid scenario based on existing regulation in Germany provides already a setting in which prosumer agents of a local energy community are economically incentivised to share their electricity with local peers and modify their consumption pattern within a client installation. Our analysis shows that the induced change in consumption behaviour has also positive side effects on the annual peaks at the network connection point of the client installation. Under a favorable scenario, this effect is even much stronger and could help to substantially reduce network congestion or compulsory curtailments, or alternatively allow more decentralised energy resources on the same grid infrastructure.

The combination of reinforcement learning for intelligent agent strategies for trading in the LEM can contribute towards converging the modelling approaches to replicate human behaviour. In addition, the ease of trading, that can be achieved with reinforcement learning have far deeper impacts in modifying existing trading approaches for administering peer-to-peer trading in different setups.

However, there are certain limitations to this simulation model. First of all, the limitation is related to data input. The household data is based on standard load curves, which although randomized through error functions, still represent an averaged electricity consumption over 15 min intervals, e.g. neglecting real existing power peaks at that level. If real load curves are obtained, a further development of various KPIs can be performed and the model can gravitate more towards reality. Also, a real load curve will provide better opportunity for load shifting since real curves have more variability amongst each other. The model helped us to test economic benefits of peer-to-peer trading in different regulatory scenarios in Germany. We identify a lack of a robust regulatory framework with clear economic advantages to explore the full potential of reinforcement learning in intelligent agent strategies and DR in LEM.

Another point to ponder upon is that we have tested one algorithm of reinforcement learning after extensive literature research. However, different LEM may have different requirements and technological and regulatory constraints. So this model is applicable for scenarios which are related in their characteristics to the particular LEM represented here and has not been proofed to be a best solution for all types of LEMs that may exist.

Conclusion and further research

The agent-based simulation model represented in this work demonstrates the application of reinforcement learning for intelligent agent strategies for peer-to-peer trading in a LEM. We have represented various regulatory scenarios and constraints with respect to German electricity regulation and showcased the opportunity of implementation of LEM in a real regulatory scenario. We have demonstrated the convergence of various strategies with changing parameters of the modified Erev-Roth algorithm, thus giving the participants flexibility to choose between different strategies with different gains and penalties. We have also demonstrated the application of DR to reduce dependency on the grid, provide economic benefit to individual agents and grid flexibility for a LEM. In addition, we have presented a sensitivity analysis of the impact of increase of renewable resources and more peak shading based on price sensitive DR in a LEM. To analyse the regulatory scenarios and provide a test bench for simulating different implications of LEM, we have set different scenarios based on the level of interaction between agents in the simulation. It is demonstrated that a degree of local sufficiency of more than 80% can be achieved with increase of renewable and DR% as demonstrated in Fig. 3. Also, a significant economic benefit for the LEM was achieved by decreasing the average price of electricity up to 8c€/kWh. The annual residual peak demand of electricity of the entire LEM was reduced even with small load shifting through price-based DR.

Further research should be targeted towards technological and policy standards of different countries of Europe and world-wide to verify the application of the model for different regulatory contexts. In this article a perfect forecast was assumed to simulate different scenarios. However, in reality, the this may not be the case. So, the study of deviation of consumption from forecasted demand and its impact on the reinforcement learning strategy is an interesting paradigm that must be further investigated. Also, significant research is developing towards Q-learning algorithms and deep reinforcement learning for application in LEMs which should be further explored. The pricing strategy of the model is based on reinforcement learning which targets to decrease the MCP in times of high generation and increase the price during time intervals of high consumption. However, the pricing strategy does not incorporate any price for congestion in the network as high generation often leads to network congestion in real world scenario and this point should be further investigated. In addition, our reinforcement learning approach is focused purely on achieving economic benefits, whereas real world scenarios can have broader inclusion of other benefits (i.e. achieve energy independence for communities, provide substantial grid flexibility, increase of renewables in the total energy mix etc.). We also recommend to include the non-economic objectives of LEMs in future research.

Abbreviations
ABM: Agent-based modelling; CAES: Consumer Automated Energy Management System; CHP: Combined heat-and-power; DER: Distributed Energy Resources; DLS: Degree of local sufficiency; DR: Demand response; DSM: Demand side management; EC: European Commission; EE: Energy efficiency; KPI: Key performance indicators; LEM: Local

energy market; MCP: Market closing price; PV: Photovoltaic; RES: Renewable energy sources; RPD: Residual peak demand; SDR: Shaded peak of Demand Response; SR: Spinning reserve; USD: United States Dollar; TOU: Time-of-use tariff

Availability of data and materials
The data-sets used and/or analyzed during the current study are available from the corresponding author on reasonable request. The PV generation data is not publicly available as it is self-measured over the time of several projects. mCHP data used in this study originates from field tests in cooperation with EDF R&D and is therefore not publicly available. The household demand data is based upon the standardized data sets by Unna (2002).

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]European Institute for Energy Research, Emmy-Noether-Str. 11, 76131 Karlsruhe, Germany. [2]MK Consulting, Hildebrandtstr. 5, 76227 Karlsruhe, Germany. [3]Karlsruhe Institute of Technology, Fritz-Erler-Str. 23, 76131 Karlsruhe, Germany.

References
Albadi MH, El-Saadany EF (2008) A summary of demand response in electricity markets. Electr Power Syst Res 78(11):1989–1996
Bishop CM (2006) Pattern recognition and machine learning. Springer, New York
Caramizaru A, Uihlein A (2020) Energy Communities: An Overview of Energy and Social Innovation. Publ Off Eur Union Luxemb EUR 30083 EN:7–11
Chen T, Alsafasfeh Q, Pourbabak H, Su W (2018) The next-generation us retail electricity market with customers and prosumers *A bibliographical survey*. Energies 11(1):8
Chen T, Bu S (2019) Realistic Peer-to-Peer Energy Trading Model for Microgrids using Deep Reinforcement Learning. In: 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe). IEEE, Chengdu. pp 1–5
Chen T, Su W (2018a) Local energy trading behavior modeling with deep reinforcement learning. IEEE Access 6:62806–62814
Chen T, Su W (2018b) Indirect customer-to-customer energy trading with reinforcement learning. IEEE Trans Smart Grid 10(4):4338–4348
Claessens B, Vandael S, Ruelens F, De Craemer K, Beusen B (2013) Peak shaving of a heterogeneous cluster of residential flexibility carriers using reinforcement learning. In: IEEE PES ISGT Europe 2013. IEEE, Lyngby. pp 1–5
Claessens BJ, Vandael S, Ruelens F, Hommelberg M (2012) Self-learning demand side management for a heterogeneous cluster of devices with binary control actions. In: 2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe). IEEE, Berlin. pp 1–8
de Oliveira e Silva G, Hendrick P (2017) Photovoltaic self-sufficiency of belgian households using lithium-ion batteries, and its impact on the grid. Appl Energy 195:786–799
Dimeas AL, Hatziargyriou ND (2010) Multi-agent reinforcement learning for microgrids. In: IEEE PES General Meeting. IEEE, Minneapolis. pp 1–8
Erev I, Rapoport A (1998) Coordination, "magic," and reinforcement learning in a market entry game. Games Econ Behav 23(2):146–175
Erev I, Roth AE (1998) Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. Am Econ Rev 88:848–881
Guo Y, Zeman A, Li R (2009) A reinforcement learning approach to setting multi-objective goals for energy demand management. Int J Agent Technol Syst (IJATS) 1(2):55–70
Hastie T, Tibshirani R, Friedman J (2009) Unsupervised learning. In: The elements of statistical learning. Springer, New York. pp 485–585
International Energy Agency (2018) World Energy Investment 2018. Int Energy Agency. https://webstore.iea.org/world-energy-investment-2018. Accessed 20 Oct 2019

Jensen RH, Strengers Y, Kjeldskov J, Nicholls L, Skov MB (2018) Designing the Desirable Smart Home: A Study of Household Experiences and Energy Consumption Impacts. Association for Computing Machinery, New York. https://doi.org/10.1145/3173574.3173578

Kaelbling LP, Littman ML, Moore AW (1996) Reinforcement learning: A survey. J Artif Intell Res 4:237–285

Koirala BP, Koliou E, Friege J, Hakvoort RA, Herder PM (2016) Energetic communities for community energy: A review of key issues and trends shaping integrated community energy systems. Renew Sust Energ Rev 56:722–744

Koza JR, Bennett FH, Andre D, Keane MA (1996) Automated design of both the topology and sizing of analog electrical circuits using genetic programming. In: Artificial Intelligence in Design'96. Springer, Dordrecht. pp 151–170

Kuznetsova E, Li Y-F, Ruiz C, Zio E, Ault G, Bell K (2013) Reinforcement learning for microgrid energy management. Energy 59:133–146

Long C, Wu J, Zhou Y, Jenkins N (2018) Peer-to-peer energy sharing through a two-stage aggregated battery control in a community microgrid. Appl Energy 226:261–276

Maringer D, Ramtohul T (2012) Regime-switching recurrent reinforcement learning for investment decision making. Comput Manag Sci 9(1):89–107

Marzband M, Ghadimi M, Sumper A, Domínguez-García JL (2014) Experimental validation of a real-time energy management system using multi-period gravitational search algorithm for microgrids in islanded mode. Appl Energy 128:164–174

Marzband M, Sumper A, Domínguez-García JL, Gumara-Ferret R (2013) Experimental validation of a real time energy management system for microgrids in islanded mode using a local day-ahead electricity market and minlp. Energy Convers Manag 76:314–322

Mazidi M, Zakariazadeh A, Jadid S, Siano P (2014) Integrated scheduling of renewable generation and demand response programs in a microgrid. Energy Convers Manag 86:1118–1127

Mendes G, Nylund J, Annala S, Honkapuro S, Kilkki O, Segerstam J (2018) Local energy markets: opportunities, benefits, and barriers. In: CIRED Workshop-Ljubljana, Paper. AIM Association, CIRED, Liege. pp 1–4

Mengelkamp E, Bose S, Kremers E, Eberbach J, Hoffmann B, Weinhardt C (2018a) Increasing the efficiency of local energy markets through residential demand response. Energy Inf 1(1):11

Mengelkamp E, Gärttner J, Rock K, Kessler S, Orsini L, Weinhardt C (2018b) Designing microgrid energy markets: A case study: The brooklyn microgrid. Appl Energy 210:870–880

Mengelkamp E, Gärttner J, Weinhardt C (2018c) Intelligent agent strategies for residential customers in local electricity markets. In: Proceedings of the Ninth International Conference on Future Energy Systems. Association for Computing Machinery, New York. pp 97–107

Mengelkamp E, Gärttner J, Weinhardt C (2018d) Decentralizing energy systems through local energy markets: the LAMP-project. In: Multikonferenz Wirtschaftsinformatik. Institut für Wirtschaftsinformatik, Leuphana Universität Lüneburg, Lüneburg

Moody J, Saffell M (2001) Learning to trade via direct reinforcement. IEEE Trans Neural Netw 12(4):875–889

Nanduri V, Das TK (2007) A reinforcement learning model to assess market power under auction-based energy pricing. IEEE Trans Power Syst 22(1):85–95

Nicolaisen J, Petrov V, Tesfatsion L (2001) Market power and efficiency in a computational electricity market with discriminatory double-auction pricing. IEEE Trans Evol Comput 5(5):504–523

Olivella-Rosell P, Lloret-Gallego P, Munné-Collado Í, Villafafila-Robles R, Sumper A, Ottessen S, Rajasekharan J, Bremdal B (2018) Local flexibility market design for aggregators providing multiple flexibility services at distribution network level. Energies 11(4):822

Palensky P, Dietrich D (2011) Demand side management: Demand response, intelligent energy systems, and smart loads. IEEE Trans Ind Inf 7(3):381–388

Panait L, Luke S (2005) Cooperative multi-agent learning: The state of the art. Auton Agent Multi-Agent Syst 11(3):387–434

Pilz M, Al-Fagih L (2017) Recent advances in local energy trading in the smart grid based on game-theoretic approaches. IEEE Trans Smart Grid 10(2):1363–1371

Pinson P, Madsen H, et al (2014) Benefits and challenges of electrical demand response: A critical review. Renew Sust Energ Rev 39:686–699

Saele H, Grande OS (2011) Demand response from household customers: Experiences from a pilot study in norway. IEEE Trans Smart Grid 2(1):102–109

Shariatzadeh F, Mandal P, Srivastava AK (2015) Demand response for sustainable energy systems: A review, application and implementation strategy. Renew Sust Energ Rev 45:343–350

Shimokawa T, Suzuki K, Misawa T, Okano Y (2009) Predicting investment behavior: An augmented reinforcement learning model. Neurocomputing 72(16-18):3447–3461

Siano P (2014) Demand response and smart grids—a survey. Renew Sust Energ Rev 30:461–478

Silver D (2015) Lecture 1: Introduction to Reinforcement Learning. Google DeepMind 1:1–10

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction MIT Press. MIT Press, Cambridge MA

Unna S (2002) VDEW-BDEW-Lastprofile. http://tinyurl.com/j4qa7qb. Accessed 20 June 2019

Vázquez-Canteli JR, Nagy Z (2019) Reinforcement learning for demand response: A review of algorithms and modeling techniques. Appl Energy 235:1072–1089

Weidlich A, Veit D (2008) A critical survey of agent-based wholesale electricity market models. Energy Econ 30(4):1728–1759

Wu Q, Guo J (2004) Optimal bidding strategies in electricity markets using reinforcement learning. Electr Power Components Syst 32(2):175–192

Zhou S, Zou F, Wu Z, Gu W, Hong Q, Booth C (2020) A smart community energy management scheme considering user dominated demand side response and p2p trading. Int J Electr Power Energy Syst 114:105378

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.