

A Deep Demand Response Program for Local Electricity Systems

Marie-Louise Arlt^{1,2} , Gunther Gust³  and Dirk Neumann⁴

Production and Operations Management
2026, Vol. 35(3) 1020–1037
© The Author(s) 2025



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10591478251363654
journals.sagepub.com/home/pao



Abstract

The decarbonization of power systems facilitates the electrification of appliances, many of which can be operated in a flexible way. Demand response (DR) programs can exploit this flexibility with retail price adjustment, thereby addressing several operational challenges. In this paper, we address the welfare optimization problem of local utilities that procure electricity for their customers at the wholesale market. We demonstrate how DR programs can be designed for local electricity systems where electricity demand and its response to temporary price changes is unknown. For this purpose, we address a novel and complex pricing problem—pricing under unknown, time-interdependent, and discontinuous demand—leveraging Deep Reinforcement Learning. Using a numerical case study calibrated on Californian electricity market data, we show that such a “Deep DR program” helps to identify effective prices that improve social welfare. The performance of the program is consistently positive across a variety of system conditions. We further demonstrate that our approach beats Time-of-Use tariff-based benchmarks already after five and a parametric benchmark after 19 simulation days, on average. Second, we provide novel insights regarding an important but frequently overlooked aspect of DR program design: The length of the notification interval, that is the timespan for which future prices must be set in advance. We find that the timing of price information is important and that longer notification intervals can improve social welfare. Finally, we provide insights into DR price setting and find that DR prices co-move with wholesale market prices but are lower for longer notification intervals and shorter event sequences. The presented Deep DR program provides an example of how advances in machine learning-based algorithms can help to meet the complex operational requirements of future local electricity systems.

Keywords

Electricity Markets, Demand Response, Algorithmic Pricing, Optimal Control, Deep Reinforcement Learning

Date received 13 September 2023; accepted 11 July 2025 after three revisions

Handling Editor: Vijay Mookerjee

1 Introduction

The decarbonization of the economy requires the electrification of many applications and gives rise to a large number of new loads in local electricity systems, such as electric vehicles and heat pumps. Electricity consumption of these loads is, to some degree, flexible in time. Similarly, many pre-existing loads, such as production facilities, have a large potential for flexible operations (e.g. European Commission, 2022; US Department of Energy, 2022). This temporal flexibility in local electricity systems can be exploited via demand response (DR) programs. Such programs temporarily change prices to incentivize customers to adjust their load. Thereby, operational challenges, such as short-term imbalances of supply and demand due to fluctuating renewable energies, can be addressed more effectively.

While DR programs have generally received considerable attention by the research community (Boßmann and Eser,

2016; Wu et al., 2025), this is not the case for their application in local electricity systems. The majority of contributions to research on DR programs investigate applications at

¹ Department of Law, Business, and Economics & Bavarian Center for Battery Technology (BayBatt), University of Bayreuth, Bayreuth, Germany
² ifo Institute, Munich, Germany

³ Center for Artificial Intelligence and Data Science (CAIDAS) & Faculty of Management and Economics, University of Würzburg, Würzburg, Germany

⁴ Department of Economics and Behavioural Sciences, Albert Ludwig University of Freiburg, Freiburg im Breisgau, Germany

Corresponding author:

Marie-Louise Arlt, Department of Law, Business, and Economics & Bavarian Center for Battery Technology (BayBatt), University of Bayreuth, Universitaetsstrasse 30, 95447 Bayreuth, Germany; ifo Institute, Poschingerstrasse 5, 81679 Munich, Germany.
Emails: arlt@uni-bayreuth.de; arlt@ifo.de

high levels of aggregation, such as electricity procurement on wholesale markets (Märkle-Huß et al., 2018; Yousefi et al., 2011) or reduction of peak generation (Blonz, 2016). At the local level, research into DR programs mostly targets one specific type of load, such as electric vehicles (Valogianni et al., 2020), or heating, ventilation and air conditioning (HVAC) systems (Adelman and Uçkun, 2019). In contrast, in our work, we address local electricity systems, such as electricity distribution systems, with an unknown combination of different load types.

Determining effective DR prices for local electricity systems, however, is challenging. Due to missing communication infrastructure at customer premises and privacy considerations (Haider et al., 2016), the composition of loads and their individual responses to price changes is unknown. Instead, it is only possible to observe an aggregate response. Determining effective prices when observing only aggregate load behavior is difficult. First, the loads in distribution systems exhibit different kinds of temporal flexibility and thus respond differently to price changes: For instance, during periods of high prices, some production processes can be paused, while others must not be interrupted, and local electricity storage facilities may even feed electricity back into the grid. Thus, the reaction of the system to price changes depends on the composition of the loads. Second, many electrical applications, such as production processes, have on-off properties. This introduces discontinuities into the demand function and makes the estimation of effective prices methodologically challenging.

In this paper, we overcome these challenges and thereby provide three contributions. First, we contribute to research on DR programs by demonstrating how DR programs can be designed for local electricity systems. Specifically, we address the welfare optimization problem of local utilities—such as energy cooperatives and regulated utilities in the United States or utilities owned by municipalities in Europe—which procure electricity at the wholesale market. We show that DR programs can perform effectively even if the properties of electricity demand are unknown and wholesale market prices are highly variable. The performance remains robust under a variety of system characteristics such as different load type combinations, wholesale price variability, and forecasting quality. Moreover, the underlying DR prices can be identified quickly such that welfare improvements exceed the ones achieved by a Time-of-Use (ToU) tariff, that is a pricing structure where the cost of electricity varies depending on the time of day, after a number of training steps corresponding to five days in the real-world, already. The performance of a parametric benchmark, that optimizes prices using a linear model of demand, is exceeded after 19 days. This underlines the practical applicability of the program.

As the second contribution, to design the DR program, we solve a novel, complex pricing problem: pricing under unknown, time-interdependent, and discontinuous demand. First, the response of customers to prices is *unknown* ex-ante. Second, customers' decisions depend on the state of

their loads, such as the state-of-charge of an electric vehicle or the state-of-completion of a production job. These states render the demand function *time-interdependent*, so that not only current prices but also past and future prices affect the consumption decisions (resulting in rebound and prebound effects). Third, the on-off property of devices introduces *discontinuities* into the demand function. This problem of finding prices under unknown, time-interdependent, and discontinuous demand also occurs in other contexts, such as marketing, and has not yet been solved optimally (cf., Bertsimas and Vayanos, 2017; den Boer and Keskin, 2020). We show how machine learning-enabled pricing based on Deep Reinforcement Learning (RL) is generally able to solve this problem quickly and effectively.

Third, we provide novel insights regarding an important, but frequently overlooked aspect of DR program design: The length of the notification interval, that is the timespan for which future prices must be set in advance. When participating in a DR program, load operators decide upon the dispatch of their appliances based on current and upcoming prices. Depending on the constraints of the load, such a decision may lock in demand for the upcoming hours and reduce future load flexibility. Due to such reductions in load flexibility, the length of the notification interval affects the performance of DR programs (Kienscherf et al., 2020). However, this mechanism has not yet been studied systematically. We show how effective choices of notification intervals depend on the forecasting capabilities of the DR program operator, customers' expectations, and the participating loads.

We proceed as follows. In Section 2, we explain the gaps in research on DR programs and pricing under unknown demand. In Section 3, we present the structure of the DR program analyzed by us, including the optimization problems of the DR program operator and the flexible load operators. We characterize the optimal decisions of the load operators in Section 4. In Section 5, we describe the solution of the DR program operator's decision problem using Deep RL ('Deep DR program'). In Section 6, we evaluate the performance of the Deep DR program and present results with regard to learning speed, load portfolios, and other important characteristics. We conclude and discuss our findings in Section 7.

2 Related Work

Time-variable prices are an important lever for coordinating supply and demand in electricity systems. On the supply side, the importance of pricing for investment in and operations of generation resources has been studied for a long time (e.g., Bohn et al., 1984). The advent of variable renewable energy sources, such as solar and wind energy, has made pricing choices more complex. As a result, the literature has revisited pricing questions in the context of renewable energies (e.g., Gambardella and Pahle, 2018; Sunar and Swaminathan, 2021; Zhou et al., 2016). These include, on the one hand, the impact

of renewable energies on prices and revenues of market participants (e.g., Al-Gwaiz et al., 2017; Sunar and Birge, 2019) and, on the other hand, the impact of pricing on renewable energy investments (e.g., Comello and Reichelstein, 2017; Kök et al., 2020).

Another important lever for managing more volatile electricity systems lies at the demand side. By adjusting electricity prices during critical times, demand can be incentivized to deviate from original consumption patterns, which is the purpose of DR programs.

2.1 DR Programs

DR programs are an active field of research (see e.g., Faruqui and Sergici, 2010; Haider et al., 2016; Shoreh et al., 2016; Vardakas et al., 2015; Yan et al., 2018 for reviews). DR programs are usually characterized by a temporary increase of retail prices from a baseline (US Department of Energy, 2006). The number of price increases is typically restricted by the regulator, in order to provide a price insurance for customers and limit their exposition to price risks (Borenstein, 2007). The limitation of price increases differentiates DR programs from real-time pricing (Borenstein, 2005). The effectiveness of DR programs has been studied for a variety of applications at different levels of aggregation, which we summarize in the following.

One large stream of research studies applications of DR programs at high levels of aggregation, such as national electricity systems. At this level, DR programs can help to shift load into times of low supply cost as well as reduce demand peaks and congestion. As a result, DR programs can reduce supply costs (e.g., Feuerriegel and Neumann, 2016; Gils, 2016; Märkle-Huß et al., 2018) and save on expensive investments in generation (e.g., Blonz, 2016) and transmission networks (e.g., Nikzad et al., 2012; Zhou and Mai, 2021).

Another part of the literature considers the level of individual customers. These studies investigate how DR programs should be designed for optimal load response, for instance with regard to the role of baselines (Agrawal and Yücel, 2022), suitable combinations of long-term contracts and real-time adjustment (Wang et al., 2017), or optimal contracts considering both demand reduction and volatility (Aïd et al., 2022). Webb et al. (2016) further consider the interaction between DR programs and energy efficiency measures. Based on the assumption that customer models are known to the utility, these papers are able to determine the optimal DR price and other relevant design parameters.

Our work lies between these two extremes, since we consider groups of customers. Other contributions at this level include Anees et al. (2021), Grimm et al. (2021), Liang et al. (2019), and Sarfarazi et al. (2022). In these studies, the DR program operator faces a group of available flexible loads but their constraints are known. Therefore, the authors are able to identify the effective price by solving a complex optimization problem, either iteratively or by using linear programming. In

our study, however, we are interested in situations in which the population of customers and their constraints are unknown to the DR program operator and their individual responses are not observable.

Few papers also consider groups of customers and provide approaches to designing DR under such less defined conditions. Adelman and Uçkun (2019) design a social welfare-maximizing program for a population of automated HVAC systems while Valogianni et al. (2020) address electric vehicles for a system under grid constraints. In both cases, the DR program operator requires an algorithm to approximate the optimal price but the programs address homogeneous groups of loads. As a result, the operator knows the structure of the customers' operational problem. However, in local electricity systems, there are heterogeneous electric load types (e.g. HVAC systems; electric storages; electric vehicles; interruptible and noninterruptible loads), which have different degrees of flexibility and thus respond differently to prices. DR programs to manage such local electricity systems with heterogeneous loads have not been researched. We follow Adelman and Uçkun (2019) and Valogianni et al. (2020)'s setup of designing a DR program for a population of customers with unknown preferences. In addition, we assume that the DR program operator does not know the demand structure of the customers but instead faces an unknown portfolio of heterogeneous load types in the system. Accordingly, a key methodological challenge within the design of the local DR program is the identification of prices under unknown demand.

2.2 Pricing Under Unknown Demand

Vardakas et al. (2015) review optimization approaches for DR programs where customer preferences and operational constraints are known. Under such conditions, optimal DR prices can be identified by solving a linear programming problem (e.g., Anees et al., 2021; Grimm et al., 2021) or iteratively (e.g., Sarfarazi et al., 2022). However, DR program operators usually only have limited knowledge of the existing load flexibility of their portfolio - due to a lack of communication infrastructure and privacy considerations (Haider et al., 2016). Moreover, the pricing problem faced in local DR programs is particularly challenging as demand can be time-interdependent and discontinuous. The demand function is therefore a piece-wise function, for which both the limits of the subdomains as well as the functional form on each subdomain are unknown.

In the context of DR programs and electricity markets, several suggestions have been made to identify effective prices under unknown demand. In most cases, the existing literature approaches the problem by assuming a reasonable parametric form of demand—usually a demand function linear in price—which is subsequently parameterized. For the problem of DR pricing for continuous demand with no time-interdependency, Khezeli et al. (2017) propose least squares and quantile estimation for optimizing incentive payments in a DR program. They define a linear demand function, derive

a truncated least squares estimator, and evaluate a myopic and a perturbed myopic price policy for DR contracts. For a similar demand model and solution approach, Mieth and Dvorkin (2020) suggest a pricing strategy for a distribution grid with grid constraints. Other approaches additionally assume time-interdependency of demand. Dong et al. (2017) empirically estimate a cross-price elasticity between peak and off-peak periods and then analytically derive optimal ToU tariffs. Nikzad et al. (2012) instead numerically solve a stochastic linear model to optimize grid reliability, although for known system parameters. However, the suggested approaches to DR price estimation cannot be easily transferred to a local electricity system where demand can be discontinuous.

Other contributions have therefore suggested parameterized pricing approaches that go beyond linear demand functions. Adelman and Uçkun (2019) and Valogianni et al. (2020) suggest adaptive pricing algorithms under unknown preferences for major appliances in the distribution grid—HVAC systems and electric vehicles. However, their approaches are specific to the appliances considered as they leverage the operational constraints of these appliances. Such model-based learning approaches have also been used in other fields where agents face unknown demand. For example, Cohen et al. (2020) suggest a multidimensional version of binary search to approximate the optimal price of products whose market value linearly depends on a set of features, such as those found in online flash sales. Moazeni et al. (2020) optimize marketing efforts in a multiplicative advertising exposure model with Poisson arrivals.

As assumptions of parametric models on demand are strong and may lead to inconsistent results, nonparametric approaches can help to alleviate the restrictions imposed by a target demand model. Vallés et al. (2018) use quantile regressions to derive distribution functions of flexibility conditional on socio-economic and physical parameters, allowing them to classify multiple customer groups with different response behavior. However, they do not consider time-interdependencies. As a hybrid of parametric and nonparametric approaches, Yousefi et al. (2011) consider that electricity demand can be composed of different types of aggregate demand functions and suggest a Q learning-based approach to identifying the weights and parameters of a linear combination of four demand models. More generally, Harsha et al. (2021) further develop a quantile-based method to address the newsvendor problem and apply it to energy consumption data of residential households. Such an approach allows for a wider variety of demand functions as it does not require assumptions about the specific shape of demand. However, while they consider past demand as a potential parameter in their model, their approach does not allow for inter-temporal optimization.

RL is one specific approach to optimize prices in a nonparametric way while considering time-interdependencies. Recently, RL has been used to address a variety of operational problems in the smart grid (e.g., Li et al., 2023; Ponce et al.,

2024; Vázquez-Canteli and Nagy, 2019; Zhang et al., 2020). In the context of DR programs, these include the control problem of consumer loads (e.g., Bahrami et al., 2021), as well as the price setting problem. While this literature on pricing in DR considers differently detailed load models—from the retail pricing problem (e.g., Lu et al., 2018; Qian et al., 2024; Tang et al., 2024), via the joint pricing and bidding problem at retail and wholesale markets (e.g., Xu et al., 2019, 2022b), to the consideration of market power in wholesale market bidding (Xu et al., 2024)—its models of the consumer side are often simplified and do not incorporate key aspects of load flexibility: The on-off property of devices is ignored and demand flexibility follows the same continuous functional form for all consumers (e.g., Kim et al., 2015; Liu et al., 2021; Xu et al., 2022a). While the functions commonly take into account price elasticities (Liu et al., 2021), that may also be time-dependent and vary over the course of the day (Lu et al., 2018), storability, shiftability and interruptibility are commonly not modeled, even though these are key properties of flexible load (Barth et al., 2018)—and also pose methodological challenges to pricing. Moreover, all these contributions assume homogeneous customer models, such that the behavior of a consumer portfolio composed of several types of flexibility has not been explored.

To summarize, we extend the literature on DR design by considering more complex and heterogeneous types of consumer flexibility. In particular, we investigate settings where the DR program operator interacts with a portfolio of consumers of unknown flexibility types, whose behavior varies over the course of the day and whose load response exhibits time-interdependencies and discontinuities.

By providing an effective way to solve pricing problems under unknown, time-interdependent, and discontinuous demand, our approach also relates to the theoretical literature on pricing under unknown demand (see e.g. den Boer, 2015, for an overview). Most notable contributions include Bertsimas and Vayanos (2017), who suggest an adaptive optimization approach for price setting under a linear demand model with extensions for time-interdependent or polynomial demand; den Boer and Zwart (2015), who generalize learning for non-time-interdependent demand; and den Boer and Keskin (2020), who consider discontinuous demand functions. In contrast to these contributions, our setting is more general, since it is characterized by both time-interdependencies and discontinuities. Despite this complex setting, we later show that we still find effective solutions to the pricing problem.

3 The Local DR Program

In local electricity systems, small regulated utilities or publicly-owned cooperatives procure electricity on behalf of their retail customers. Part of this electricity is bought at the wholesale market, such as the Western Energy Imbalance Market in the case of California (CAISO, 2023b). While prices at

the wholesale market vary in real-time, utilities and cooperatives typically sell the electricity to their customers at a retail tariff that is either entirely constant or takes multiple price levels, depending on the hour of the day (so called “Time-of-Use” tariffs). Given the monopolistic position of these utilities, retail tariffs are subject to the approval of the regulator or the members of the cooperative (e.g., California, 2021; CPUC, 2019).

Customers in local electricity systems are residential households as well as small commercial and industrial customers. The roll-out of smart meters is still limited in some countries (pwc, 2022). In this case, utilities cannot observe individual customers’ response to time-specific prices in real-time but only *ex post* when billing their customers (typically at the end of the calendar month or year). The utilities and cooperatives can, however, measure electricity consumption in their local electricity system in real-time, e.g. through information and communication technology-equipped metering infrastructure at transformer stations.

We now assume that utilities and cooperatives are price takers at the wholesale market and operate a DR program (and in the following, we refer to them as “DR program operators”). For instance, the peak load of Palo Alto Utilities, an example for such a utility and potential DR program operator, is 159 MW as compared to the overall peak load of 52 GW in California/CAISO (CAISO, 2023a; ORNL, 2023). Because of these differences in magnitude, we consider the modeling as price takers reasonable. Under the DR program, the operator can temporarily raise retail prices during periods of high wholesale market prices and thereby incentivize its customers (in the following, we refer to them as “load operators”) to shift consumption to periods of lower wholesale market prices. As a result, electricity procurement costs at the wholesale market can be decreased. The price increases are, however, only permitted during times of high wholesale market prices, in order to limit the exposition of customers to price risks. Such limiting conditions for price deviations are common in related DR programs, such as the Critical Peak Pricing (CPP) program in California (Blonz, 2016), for instance.

3.1 Temporal Structure of the DR Program

Figure 1 illustrates the temporal structure of the DR program and the sequential decision-making of the DR program operator and the load operators: In the first stage of time period t , the DR program operator makes a forecast with regard to the upcoming wholesale market price in period $t+n$. If, in period t , higher than usual wholesale market prices are forecasted for period $t+n$, the DR program operator can choose to deviate from the base retail price p^b and set a DR price $p_{t+n} > p^b$ instead. We refer to n as the *notification interval*, with $n \in \mathbb{N}$. Retail prices \bar{p}_t for periods $\{t, t+1, \dots, t+n\}$ are known to load operators at time t ; retail prices beyond $t+n$ are uncertain. An overview of the notation used throughout the paper can be found in Supplement A.

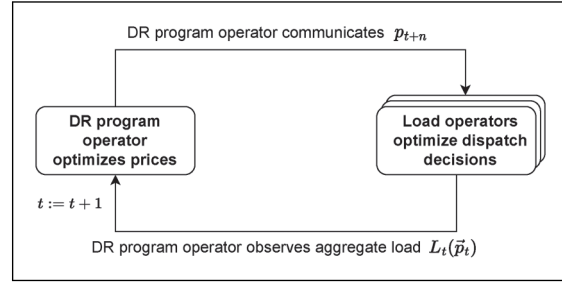


Figure 1. Sequence of actions in period t of the demand response (DR) program.

In the second stage of the time period t , the cost-minimizing load operators observe p_{t+n} , re-optimize their load schedule for expected operations costs, and implement their dispatch scheduled for t . At the end of period t , the DR program operator observes the resulting aggregated load $L_t(\bar{p}_t)$ in the local electricity system and updates his beliefs about the aggregate demand function as well as the optimal DR price-setting policy. Importantly, the DR program operator knows neither the optimization problems nor the dispatch of individual load operators. The two-stage process then repeats for subsequent time steps.

3.2 Decision Problem of the DR Program Operator

Since utilities and cooperatives are owned by regional authorities and customers, we assume that the DR program operator pursues welfare maximization (cf. Adelman and Uçkun, 2019). The DR program operator seeks to maximize expected social welfare by the choice of retail prices \bar{p}_t . Social welfare is the difference between the utility of customers from consumption, $U_t(\bar{p}_t)$, and the cost of procuring electricity on the wholesale market, $p_t^{WS} \cdot L_t(\bar{p}_t)$, where p_t^{WS} corresponds to the wholesale market price. Equation (1) formalizes the decision problem of the DR program operator,

$$\begin{aligned} \max_{\mathbf{p}} \quad & \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left(\sum_t \{U_t(\bar{p}_t) - p_t^{WS} \cdot L_t(\bar{p}_t)\} \right) \quad (1) \\ \text{s.t.} \quad & p_t = \begin{cases} \in \mathbf{P}^{DR}, & \forall t \in \mathbf{T}^{DR}, \\ p^b & \forall t \notin \mathbf{T}^{DR}. \end{cases} \quad (2) \end{aligned}$$

The DR program operator’s pricing is subject to two constraints. First, a DR event can only be called if the forecasted wholesale market price $\mathbb{E}(p_t^{WS})$ exceeds a certain threshold \bar{p}^{WS} , as formalized in Equation (2). Second, Equation (2) requires that the DR price falls into a set of possible DR prices \mathbf{P}^{DR} (e.g. defined by a maximum price). Both constraints reflect real-world regulatory constraints that ensure that load operators are protected from excessive price increases.

The objective function as described in Equation (1) cannot be operationalized directly, since the utility $U_t(\bar{p}_t)$ of load

operators is unknown to the DR program operator. We solve this issue as follows. First, we decompose utility $U_t(\vec{p}_t) := \bar{U}(p^{noDR}) - \Delta U_t(\vec{p}_t)$, where $\bar{U}(p^{noDR})$ represents the utility experienced from consumption under a fixed retail tariff p^{noDR} (when no DR program applies) and $\Delta U_t(\vec{p}_t)$ the utility change under the DR program. As $\bar{U}(p^{noDR})$ is a constant, the optimization problem of Equation (1) is equivalent to the following,

$$\max_{\mathbf{p}} \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left\{ \sum_t \left[\underbrace{-\Delta U_t(\vec{p}_t)}_{\text{Utility change of consumers}} - \underbrace{p_t^{WS} \cdot L_t(\vec{p}_t)}_{\text{Wholesale market procurement costs}} \right] \right\}. \quad (3)$$

Second, we approximate demand $L_t(\vec{p}_t)$ within a time period t by a linear function $\tilde{L}_t(\vec{p}_t)$. Considering this demand model, $\Delta U_t(\vec{p}_t)$ can now be approximated by the following expression,

$$-\Delta U_t(\vec{p}_t) \sim -\frac{p_t + p^{noDR}}{2} (L_t(p^{noDR}) - \tilde{L}_t(\vec{p}_t)). \quad (4)$$

The given expression estimates the welfare change at time t as the product of the price change (as an estimate of the willingness-to-pay of the load not dispatched) and the aggregate load change, both as compared to a situation with a fixed retail price p^{noDR} . The approximation in Equation (4) relies on the assumption that $L_t(p^{noDR})$ is known. We consider this reasonable since it can be derived from historical data or an aggregate load model. In Supplement B, we explain the approximation in more detail, including a graphical illustration of $\tilde{L}_t(\vec{p}_t)$.

3.3 Decision Problems of Load Operators

After defining the optimization of the DR program operator, we now focus on the load operators. We characterize the aggregate flexible load $L_t(\vec{p}_t)$ as a sum of the contributions of four elementary flexible load types: Storage, interruptible and non-interruptible loads, and elastic loads. These elementary load types represent fundamental flexibility characteristics (Barth et al., 2018).

Figure 2 illustrates the four types and their ability to respond to variable prices: Storage is able to charge and discharge and, thereby, shift load into other time periods. Examples could be a residential battery storage or an electric vehicle. Interruptible and noninterruptible loads both have predefined load profiles which they can re-schedule. While the former allows for intermittent switching on and off, the latter needs to follow a continuously running load profile once started. Examples for interruptible loads could be a washing machine or a large-scale printing job; for non-interruptible loads, examples would be an oven for baking or

a melting process in the industry. For an elastic load, such as the level of room heating or cooling, scheduling is not time-interdependent.

Operators of time-interdependent load types—that is storage as well as interruptible and noninterruptible loads—minimize expected stage-wise electricity costs while facing an intertemporal optimization problem. We generalize the objective function of a load operator with a single time-interdependent flexible load of type $k \in \{sto, inter, ninter\}$ to the following expression,

$$\min_{\mathbf{d}} \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left(\sum_t (p_t + b_t) l_t(x_t) d_t \right). \quad (5)$$

Load operation costs are driven by two components: the retail price p_t and additional dispatch costs b_t . Both are charged per unit of energy. p_t is equal for all loads participating in the DR program while b_t differs across loads. For instance, b_t can correspond to increased labor cost for shifting load. $l_t(x_t)$ is the load profile. It describes the electricity demand of a load and is dependent on the physical state of the load, such as the state-of-charge or the state of execution at t . The decisions of the load operator are represented by d_t , which describes load activity, such as dispatching ($d_t = 1$) or not dispatching ($d_t = 0$), or charging ($d_t > 0$) and discharging ($d_t < 0$). The cost-minimizing decisions depend on the expected prices \vec{p}_t .

While all load types share the same objective function, their load type-specific operational characteristics are reflected by different sets of constraints. These constraints importantly include the intertemporal transition function $f(\cdot)$, which links the state of each load over time,

$$x_{t+1} = f(x_t, d_t). \quad (6)$$

In Supplement C.1, we provide a table summarizing the load type-specific optimization problems. We explain the type-specific model components in the next section, when deriving the solutions to the load operator's decision problems.

4 Solutions of the Load Operators' Decision Problems

We describe the scheduling problem as an optimal control problem under uncertainty to capture the dynamic decision structure of the rolling time horizon. As the DR program operator sequentially communicates future prices, the load operators take new information on prices up to $t + n$ into account. Given this price information, they independently decide upon optimal load activity d_t in t . For that purpose, load operators recursively solve the Bellman equation, minimizing the sum of dispatch costs in the current and future time periods,

$$J_t(x_t) = \min_{d_t} \{ (p_t + b_t) l_t(x_t) d_t + \mathbb{E}(J_{t+1}(x_{t+1})) \}, \quad (7)$$

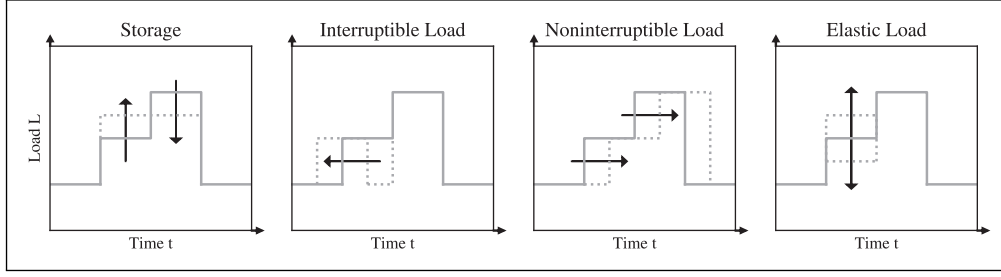


Figure 2. Flexible load types. Note. The subfigures illustrate the main flexibility characteristics of each load type. The grey line represents a customer's original load profile over time. With a flexible load, the load profile can be modified to the dashed profile. *Explanations by load type:* A customer with storage can modify his load profile by charging (increasing net load in one period) and discharging in the next period (decreasing net load in the next period). A customer with an interruptible load can anticipate part of his load, interrupt, and then exercise the rest of the load profile. In addition or alternatively, part of the load can also be postponed. A customer with a noninterruptible load can anticipate or postpone the entirety of the load profile, but not only a part of it. Finally, a customer with an elastic load can increase or decrease load in a given period in response to price, without changes to the load profile in other time periods.

where $l_t(x_t)$ is the load at state x_t , and the dispatch d_t (and thus x_{t+1}) is chosen based on current and expected prices \bar{p}_t . In the following, we characterize the optimal dispatch behavior for each load type. Proofs and a detailed description of the policies can be found in Supplements C.2 to C.5. Our results are in line with theoretical findings of the previous literature, including Zhou et al. (2016) and Barth et al. (2018).

4.1 Storage

Storage is characterized by the ability to charge electricity and discharge it at a later point in time. x_t describes the state-of-charge of the storage at the beginning of the period t . d_t denotes the charging decision by the storage operator, with $0 < d_t \leq 1$ for charging and $-1 \leq d_t < 0$ for discharging. l , a constant, represents the maximum charging and discharging rate. Storing electricity results in different types of losses. First, losses $0 < \rho^{ch} \ll 1$ occur when charging the battery ($d_t > 0$). Second, losses $0 < \rho^{dis} \ll 1$ occur when discharging the battery ($d_t < 0$). Finally, stored electricity is subject to depreciation relative to the total energy stored, defined by $0 < \rho^{loss} \ll 1$. The specification of the transition function Equation (6) reflects the coupling of the state-of-charge over time and is defined by the following piece-wise function,

$$x_{t+1} = (1 - \rho^{loss})x_t + \begin{cases} (1 - \rho^{ch})ld_t, & \text{for } d_t \geq 0; \\ \frac{1}{(1 - \rho^{dis})}ld_t, & \text{for } d_t < 0. \end{cases} \quad (8)$$

Additional constraints include the minimum and the maximum state-of-charge, x^{min} and x^{max} , that is the minimum and maximum amount of energy stored by the storage, respectively.

Let t^{dis} denote the future time period in which discharging is most profitable, excluding t . Let t^{ch} denote the future time period in which charging would be most profitable, excluding t .

PROPOSITION 4.1. *The optimal charging policy is,*

$$d_t^* = \begin{cases} \min \left\{ 1, \frac{x^{max} - x_t}{L}, d_t^{crit} \right\}, \\ \text{for } p_t < \min \{ \mathbb{E}(p_{t^{dis}}(1 - \rho^{ch})(1 - \rho^{dis}) \\ (1 - \rho^{loss})^{t^{dis} - t}), \mathbb{E}(p_{t^{ch}}(1 - \rho^{loss})^{t^{ch} - t}) \}; \\ \max \left\{ -1, -\frac{x_t - x^{min}}{L} \right\}, \\ \text{for } p_t > \mathbb{E}(p_{t^{dis}}(1 - \rho^{loss})^{t^{dis} - t}); \\ 0, \text{ else.} \end{cases} \quad (9)$$

The proof of Proposition 4.1 as well as a detailed explanation of t^{dis} and t^{ch} are given in Supplement C.2. The optimal policy distinguishes three cases: First, the storage should charge whenever (1) the price paid for charging is less than the loss-corrected price that can be recovered for discharging during the most profitable future period t^{dis} and (2) it cannot be substituted by an alternative future charging period t^{ch} for which charging would be cheaper (after correction for losses). The latter can apply if charging is profitable but the storage volume constraint binds and requires the operator to choose between different profitable charging periods. Charging occurs at the maximum amount possible, determined by either the storage rate or the available volume, unless it is not profitable to do so. In that case, charging is constrained to an internal solution of critical charging d_t^{crit} , with $0 < d_t^{crit} < 1$. In contrast, the storage should discharge whenever the current period cannot be substituted by an alternative future discharging period t^{dis} for which discharging would be more profitable (after correction for losses). Finally, the storage stays idle if none of the conditions apply. The policy described in Proposition 4.1 is similar to the heuristic provided in Zhou et al. (2019) which derives a state-dependent threshold policy for storage dispatch under positive electricity prices.

4.2 Interruptible Loads

An interruptible load follows a specified load profile $\vec{l} = (l^1, l^2, \dots, l^{|\vec{l}|})$, that is, a sequence of hourly energy consumption of length $|\vec{l}|$, that can be interrupted and restarted at a later point in time. The discrete control variable $d_t \in \{0, 1\}$ describes the activity of the interruptible load. The state x_t equals the aggregated energy consumed until the beginning of period t . It is zero at the start of the earliest time of load activity, $x_{t^{start}} = 0$, and equals the energy needed for load completion at the latest time of possible load activity, $x_{t^{end}+1} = \sum_{j=1}^{|\vec{l}|} l^j$. The transition function Equation (6) can therefore be specified as follows,

$$x_{t+1} = x_t + l(x_t)d_t, \quad (10)$$

where $l(x_t)$ is the state-dependent upcoming demand of the load vector.

Let $d^{1,*}$ denote future optimal dispatch, given a dispatch of the load in t , and $d^{0,*}$ future optimal dispatch, given no dispatch in t . Accordingly, $x^{1,*}$ and $x^{0,*}$ represent future optimal states. $(|\vec{l}| - \sum_{\tau=t^{start}}^{t-1} d_\tau)$ corresponds to the number of periods needed to finalize the load at the current state.

PROPOSITION 4.2. *The optimal dispatch policy is,*

$$d_t^* = \begin{cases} 1, & \text{for } (p_t + b_t)l(x_t) + \mathbb{E}(\sum_{\tau=t+1}^{t^{end}} \\ & \{ (p_\tau + b_\tau)(l(x_\tau^{1,*})d_\tau^{1,*} - l(x_\tau^{0,*})d_\tau^{0,*}) \}) \leq 0 \text{ or} \\ & t = t^{end} - (|\vec{l}| - \sum_{\tau=t^{start}}^{t-1} d_\tau) + 1; \\ 0, & \text{for } (p_t + b_t)l(x_t) + \mathbb{E}(\sum_{\tau=t+1}^{t^{end}} \\ & \{ (p_\tau + b_\tau)(l(x_\tau^{1,*})d_\tau^{1,*} - l(x_\tau^{0,*})d_\tau^{0,*}) \}) > 0. \end{cases} \quad (11)$$

The proof can be found in Supplement C.3. The policy distinguishes two cases. First, the next load component of the interruptible load will be dispatched in t if the expected cost change is positive as compared to being postponed. The expected cost change corresponds to the sum of net cost weighted load change from postponing. The load will also be dispatched if it is required to fulfill the load profile and comply with the terminal condition. Second, if the expected cost of dispatching in t is higher than in the future, the dispatch will get delayed.

4.3 Noninterruptible Loads

In contrast to interruptible loads, noninterruptible loads cannot be stopped once they have been started. The definition of variables and parameters as well as the transition function Equation (6) of the noninterruptible load is identical to the previously described interruptible load. In addition, the constraint

set includes an additional noninterruptibility constraint,

$$d_t \leq d_{t+1} + \frac{x_t}{\sum_{j=1}^{|\vec{l}|} l^j}. \quad (12)$$

Let t^{md} be the marginal period of dispatch, that is the time period in the future at which the noninterruptible load would best be started if not dispatched in t .

PROPOSITION 4.3. *The optimal dispatch policy for noninterruptible loads is,*

$$d_t^* = \begin{cases} 1, & \text{for } \mathbb{E}(\sum_{i=0}^{|\vec{l}|-1} [(p_{t+i} + b_{t+i}) - (p_{t^{md}+i} + b_{t^{md}+i})]l^i) \leq 0 \\ & \text{or } t = t^{end} - (|\vec{l}| - \sum_{\tau=t^{start}}^{t-1} d_\tau) + 1; \\ 0, & \text{for } \mathbb{E}(\sum_{i=0}^{|\vec{l}|-1} [(p_{t+i} + b_{t+i}) - (p_{t^{md}+i} + b_{t^{md}+i})]l^i) > 0. \end{cases} \quad (13)$$

This is a special case of Proposition 4.2. The proof can be found in Supplement C.4. In contrast to interruptible loads, noninterruptible loads cannot be interrupted and a dispatch in t determines future dispatch $d^{1,*}$, which simplifies the policy.

4.4 Elastic Loads

In addition to the time-interdependent loads, we consider elastic loads. Elastic loads respond only to the price signal of period t and do not exhibit temporal interdependencies. We choose a representation where the response is linear in price, following the example of Mieth and Dvorkin (2020). The dispatch d_t represents the scaling factor of load $l_t(p^{noDR})$ at the fixed retail price.

DEFINITION 4.1. *The dispatch policy of an elastic load is,*

$$d_t^* = \begin{cases} 0, & \text{for } \frac{p_t - p^{noDR}}{p^{noDR}} \geq \epsilon; \\ 1 - \frac{(p_t - p^{noDR})/p^{noDR}}{\epsilon}, & \text{for } \epsilon \geq \frac{p_t - p^{noDR}}{p^{noDR}}. \end{cases} \quad (14)$$

The price sensitivity parameter ϵ describes the relative price increase for which the load is reduced to zero (Gils, 2014). The changes in load are associated with a consumer surplus increase (decrease) for a load increase (decrease). Further details can be found in Supplement C.5.

4.5 Aggregate Load Response

The aggregate load behavior $L_t(\vec{x}_t, \vec{p}_t)$ corresponds to the sum of the cost-minimizing dispatch decisions $d_{j,t}^*$ of load operators, given their states \vec{x}_t and retail prices \vec{p}_t , over all loads \mathbb{L}^{sto} , \mathbb{L}^{inter} , \mathbb{L}^{ninter} , and \mathbb{L}^{el} of types storage, interruptible and

noninterruptible loads, as well as elastic loads,

$$L_t(\vec{x}_t, \vec{p}_t) = \sum_{j \in \mathbb{L}^{sto} \cup \mathbb{L}^{inter} \cup \mathbb{L}^{ninter} \cup \mathbb{L}^{el}} l_{j,t}(x_{j,t}) d_{j,t}^*(x_{j,t}, \vec{p}_t) + \sum_{j \in \mathbb{L}^{inel}} l_{j,t}. \quad (15)$$

In addition, aggregate load includes inelastic loads \mathbb{L}^{inel} , which are neither state-dependent nor responsive to the DR program. This represents nonprice responsive load as well as behind-the-meter generation resources, such as solar energy.

We can use Equation (15) to derive some simple comparative statics: *Ceteris paribus*, if the price in $t+n$ increases, load in $t+n$ weakly decreases, that is $\partial L_{t+n}(\vec{x}_t, \vec{p}_t) / \partial p_{t+n} \leq 0$. The impact on load remains, however, unclear for simultaneous price changes in adjacent periods, e.g. $\partial L_{t+n}(\vec{x}_t, \vec{p}_t) / \partial p_{t+n+1}$. For instance, an elastic load would reduce demand in $t+n$ for an increased price in $t+n$ but is not affected by the price change in $t+n+1$. Storage, however, would increase demand in $t+n$ if the price in $t+n+1$ is sufficiently larger than the price in $t+n$ and charging is profitable. This unclear response to simultaneous price changes can be explained by the characteristics of the intertemporal optimization problem of load operators.

Finally, as pricing is linear and only depends on the amount of energy consumed, the optimal dispatch of each load is independent of whether load operators hold a single load or combinations thereof. In both cases, aggregate load and the response to the DR program will be identical.

5 Solution of the DR Program Operator's Decision Problem Using Deep RL

Given the responses of the load operators described in the previous section, the DR program operator has to optimize prices (recall Equation (1)). This corresponds to an optimal control problem under uncertainty with the following unknowns: First, the upcoming wholesale market price is nondeterministic and may deviate from the DR program operator's forecast. Second, the DR program operator does not know the dispatch optimization problems of the individual load operators and cannot compute their load response to DR prices. After the DR price has been announced, the DR program operator only observes the actual aggregate load realization.

Moreover, the dispatch behavior of a single load operator can be described as a Markov decision process: the optimal decision to dispatch in t solely depends on the current state of the load x_t and knowledge in t about (expected) future prices. Consequently, the aggregated system also follows a Markov decision process as it can be derived as a summation of individual load operators' behavior (as previously stated in Equation (15)). The full system state in t can then be characterized by the state vector of all participating loads, \vec{x}_t . As the full system state is, however, unobservable for the DR program operator, we therefore propose a model-free Deep RL approach which learns and proposes effective actions for an observable state, as described in the following section.

5.1 Definition of States, Actions, and Rewards

Deep RL aims to approximate an optimal policy function—that is a function which maps the optimal action to a given state—by systematic exploration of an environment. For this purpose, at each time step, the RL agent observes the state of the system, picks an action, and considers the reward to continuously update the estimated value of an action in a given state and improve the policy.

For our setup, we choose the following specifications. We characterize the *state* s_t by the hour of the day $t\%24$ and the expected wholesale market price $\mathbb{E}(p_{t+n}^{WS})$, as forecasted in t . The hour of the day accounts for the fact that loads in electricity systems typically follow a daily seasonality. Importantly, the state does not include individual load states \vec{x}_t , as they are not observable by the DR program operator (see Section 3.1). However, the load states are a function of the seasonal dispatch patterns of loads (approximated by the hour of the day) as well as historical price choices by the DR program operator. Therefore, load states will be learned implicitly by the RL agent. The *actions* of the RL agent correspond to the retail prices p_{t+n} picked by the DR program operator. If the forecasted wholesale market price is equal to or exceeds the program-specific price threshold \bar{p}^{WS} , the DR program operator can increase the price (see Equation (2)). If that is not the case, the price remains at the base price level. As *rewards* r_t , we use the approximation of stage-wise social welfare changes defined by Equation (3),

$$r_t = \underbrace{-\frac{p_t^{DR} + p_t^{noDR}}{2} \left(L_t(p_t^{noDR}) - L_t(\vec{p}_t) \right)}_{\text{Utility change of consumers}} - \underbrace{p_t^{WS} L_t(\vec{p}_t)}_{\text{Wholesale market procurement cost}}. \quad (16)$$

Finally, the DR program operator maximizes social welfare over time, as represented by the objective function Equation (1). In Deep RL, the original objective function is substituted by the value function Q . $Q(s_t)$ corresponds to the expected value of the system state s_t , that is the discounted sum of expected future rewards for optimally chosen actions. We update our estimate of $Q(s_t)$ using the sum of the realized reward r_t and the discounted expected value of the following system state $Q(s_{t+1})$, as provided by our estimate of the value function,

$$Q(s_t) = r_t + \gamma Q(s_{t+1}). \quad (17)$$

The discount factor is described by γ . More information on the definition of value functions can be found in the work of Sutton and Barto (2020: Chapter 3.6).

5.2 Deep Deterministic Policy Gradient

The field of Deep RL has developed a variety of different algorithms for approximating optimal policies. In this paper,

we use the Deep Deterministic Policy Gradient (DDPG) algorithm as proposed by Lillicrap et al. (2015). DDPG is a policy gradient method and uses two neural networks to represent the policy (actor) and the value functions (critic). The actor network corresponds to a function mapping optimal actions to states. At each step, the actor network is used to predict the presumably optimal action, the DR price. This or a modified action (during exploration) is then implemented and the system subsequently produces a reward and transitions into the next state. At the end of the time step, the tuple of state, action, reward, and next state is used to re-train the value function and the gradients of the value function are used to update the policy.

We base our implementation of the DDPG on the parameterization suggested by Lillicrap et al. (2015) and adjust selected hyperparameters. Further details can be found in Supplement D.

6 Numerical Experiments

In this section, we evaluate the performance of the DR program that we parametrized using Deep RL (Deep DR program).

6.1 Setup

6.1.1 Wholesale Market Prices. First, we represent the wholesale market using an auto-regressive time series model with a seasonal (hourly) component and a noise term of variance σ_{res}^2 . To calibrate the default model, we use electricity prices of the Californian wholesale market (as experienced at the distribution node GLENWOOD_6_N005, for different months of the year 2020 (CAISO, 2022)). We specifically select a set of months with different degrees of price fluctuations, ranging from low fluctuations, as in traditional electricity markets, to high fluctuations, as expected for markets dominated by renewable energies.¹ In Supplement E.3, we compare the price fluctuations of the selected months to other electricity markets and show that they are comparable to markets around the world. We deploy the model to simulate different instances of the wholesale market and define the four hours with highest average prices as peak hours. More information on the estimation of the time series model can be found in Supplement E.1.

6.1.2 Forecasts of Wholesale Market Prices. For notification periods n of one hour or longer, the DR program operator relies on forecasts of wholesale market prices for the decision making on DR prices. In practice, such forecasts can be procured from specialized, third-party companies (such as AleaSoft (2024) or DNV (2024)). As forecasting errors can be temporally correlated, we model the forecasting error err_t by an AR(1) process with a white noise of variance σ_{err}^2 of 2 in the default case. The final price forecast $\mathbb{E}_{t+n}(p_t^{WS})$ at time t corresponds to the sum of the actual wholesale market price and the forecasting error. For the full mathematical model of forecasting quality, see Supplement E.2.

Instead of endogenously modeling forecasts—whose accuracy depends on local characteristics of the electricity system (such as the stability and predictability of weather conditions)—modeling synthetic forecasting errors allows to generate insights beyond specific local conditions. For instance, we later model the deterioration of the forecasting quality in the notification interval by increasing the variance σ_{err}^2 of the error term of the AR(1) process.

6.1.3 Retail Prices. We define the base price p^b experienced by participants of the Deep DR program equal to the average wholesale market price during off-peak hours. In the event that forecasted wholesale market prices increase above a threshold \bar{p}^{WS} of the off-peak average, a DR event occurs, that is, the DR program operator can deviate from the base price and set a price $p_t \in \{100\%, \dots, 500\% \} \cdot p^b$ to incentivize load adjustments.

6.1.4 Frequency of DR Events. The frequency of DR events depends on the choice of the threshold \bar{p}^{WS} . We later evaluate different thresholds between 105% and 500% of the off-peak average to resemble different real-world implementations of DR programs. A high threshold, such as $\bar{p}^{WS} = 500\%$, allows the DR operator to adjust prices only a very limited number of times, as observed in CPP programs. These programs are designed to mitigate extreme scenarios, such as heat waves, and typically allow only for rare DR events (e.g. Blonz, 2016). Conversely, a small threshold, such as $\bar{p}^{WS} = 105\%$, allows for frequent prices adjustments, as observed in variable peak pricing programs. These programs aim for a more efficient use of electricity and have daily DR events (e.g. Borenstein, 2005, 2007). In Section 6.2, we strike a middle ground and report the results for $\bar{p}^{WS} = 110\%$. The results for other thresholds are summarized in Supplement F.7.

6.1.5 Load Data. Electricity systems typically exhibit a pronounced seasonality. In our study, we use average hourly Californian demand data as the basis for the parametrization of all four load types (i.e., aggregate demand data of the entire PG&E service territory, for different months of the year 2020 (CAISO, 2022)).

The procedure for creating the load portfolio is illustrated in Figure 3 and described as follows. For *elastic loads*, we use the PG&E 24-hours load profile directly and randomly assign price sensitivities ϵ . For *interruptible* and *noninterruptible loads*, we randomly break up aggregate demand into shorter load blocks of two to five hours (dashed lines, numbered blocks). Each block represents a load profile \vec{l} during its core activity period (as illustrated in dark gray for sample load block 4). Additionally, we assign a symmetrically distributed flexibility window (arrows and areas shaded in light gray) during which randomly drawn additional dispatch costs b_t apply. Finally, the groups of interruptible, noninterruptible, and elastic loads are scaled to three equally sized groups and combined to reach 100 MW during the hour of intraday

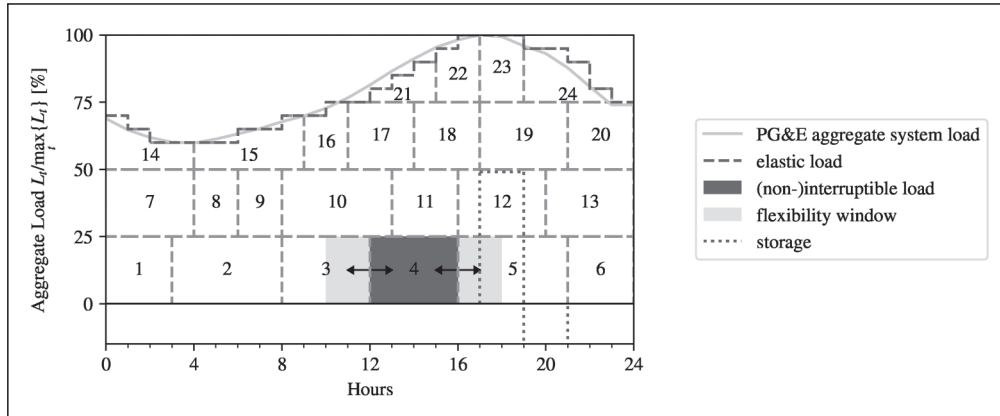


Figure 3. Illustration of the load portfolio for August 2020.

load peak at the fixed retail rate (CPUC, 2023). For *storage*, we randomly parametrize storage units according to typical volume-charging rate ratios $x^{max} : l$ and degradation losses ρ^{loss} . We assume the aggregate charging rate to equal 50% of the aggregate peak load (dotted line). In summary, this setup results in 100 loads across the four types.

Further details on load composition, the calibration of load types, as well as individual load parametrizations are presented in Supplements E.4 and E.6. Furthermore, we present a study of the behavior of the different load types when exposed to the DR program in Supplement E.5.

6.1.6 Training and Evaluation. For each parameter configuration, we perform 20 runs. For each run, we stochastically generate a new training and test time series and randomly initialize the Deep RL algorithm. We first train our DR pricing policy on a training set of 2,160 training time steps (corresponding to 2,160 hours or approximately three months in the real-world) which was generated following the procedure described in Section 6.1.1. During training, the pricing policy is optimized using the DDPG algorithm described in Section 5.

We then apply the resulting pricing policy to a test time series of 2,160 training time steps (corresponding to 2,160 hours or approximately three months in the real-world), which was also generated using the procedure described in Section 6.1.1. We evaluate the improvement in social welfare as compared to a fixed retail price p^{noDR} . The fixed retail price corresponds to the demand-weighted average wholesale market price and is about 33 % higher than the base retail price p^b of the Deep DR program.

6.1.7 Benchmarks. We further compare the results of the Deep DR program to the following three benchmarks: A parametric benchmark as well as an *ex ante* and an *ex post* ToU tariff. For the parametric benchmark, we estimate a time-of-the-day (hour)-specific demand function where demand in t depends on the current and the four previous prices. We then include this demand function into the social welfare optimization problem, which yields a convex program that we can

subsequently solve using standard optimization tools. For the ToU benchmarks, an optimized constant peak price is applied during the peak hours while the price outside these hours corresponds to the baseline price 1.0, as in the Deep DR program. For the *ex post* ToU tariff, we use the full test set in a grid search to find the peak price that yields the highest welfare. This *ex post* procedure assumes perfect foresight of future outcomes—information that is unavailable in a practical setting—and thus serves as a “best-case” or theoretical benchmark, illustrating how well a simple ToU tariff could perform under ideal conditions. For the *ex ante* ToU tariff, by contrast, we determine the optimal peak price using only the training set, reflecting the real-world constraint of setting tariffs without knowledge of future consumption patterns. Further details on the computation of the benchmarks can be found in Supplement E.7.

6.2 Results

6.2.1 Cost-Effectiveness of the Deep DR Program. Using the previously described parametrization as a starting point, Section 4 illustrates the results for each pricing approach for different months of the year, as compared to a fixed retail rate. We specifically selected these months to represent different degrees of wholesale market price fluctuations (defined as the difference between the 5 and the 95-percentiles of monthly price distributions), as they affect the potential for social welfare improvement from time-dependent load adjustments.

We find that the Deep DR program outperforms all benchmarks for the majority of the months evaluated. In the month with highest price fluctuations, August, the Deep DR approach achieves the largest advantages, achieving social welfare improvements of 382 kUSD on average, as compared to 340 kUSD (parametric), 253 kUSD (*ex post* ToU tariff), and 110 kUSD (*ex ante* ToU tariff), respectively. The performance of the Deep DR program corresponds to a 12.5% improvement over the parametric benchmark. For December and April, our approach achieves savings of 149 kUSD (versus 121 kUSD, 79 kUSD, and 83 kUSD under the benchmarks)

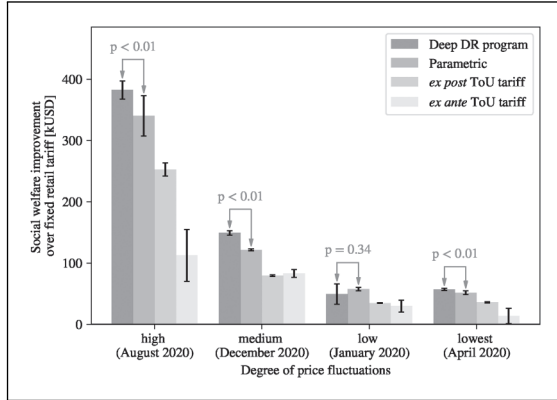


Figure 4. Comparison of pricing approaches for selected months of the year. Note. Performance (averages over 20 runs) of the Deep demand response (DR) program for different levels of wholesale market price fluctuations. Bars represent the 95% confidence interval. p values are indicated for the null hypothesis that the Deep DR program performs equally well as the parametric benchmark.

and 57 kUSD (versus 51 kUSD, 36 kUSD, and 13 kUSD under the benchmarks). Relative savings compared to the parametric benchmark are 22.8% and 10.8%, respectively. For the month of January, the social welfare improvement by the Deep DR program is 49 kUSD but it does not significantly differ from the parametric benchmark. In an additional analysis, we further systematically vary price volatility around the seasonal components and find that the social welfare improvement increases in price volatility (see Supplement F.5.1).

In summary, while the magnitude of the advantage varies, the results clearly show the superior performance of the Deep DR program. Our estimation of social welfare changes during operations aligns well with actual welfare changes (see Supplement F.1). Our results are further robust to different specifications of the setup, such as the load portfolio, forecasting quality, noise in inelastic load, the choice of the price threshold, and load operators’ expectations (see Supplements F.3 to F.8), and other system characteristics (see following sections).

6.2.2 Stability and Speed of Training. We furthermore analyze the learning behavior of the Deep RL algorithm since for real-world applicability, convergence and convergence speed are important. Section 5 illustrates the learning behavior of the algorithm for 20 runs, using the default parameters of the case study for the month of August with a storage share of 50%.

We find that, on average, we can achieve persistent savings across different runs of the stochastic price time series, as represented by the bold line. On average, the two ToU benchmarks are beaten after 120 training time steps or five days in the real-world (*ex ante* ToU tariff) and 18 days (*ex post* ToU tariff), respectively. The parametric benchmark is beaten after

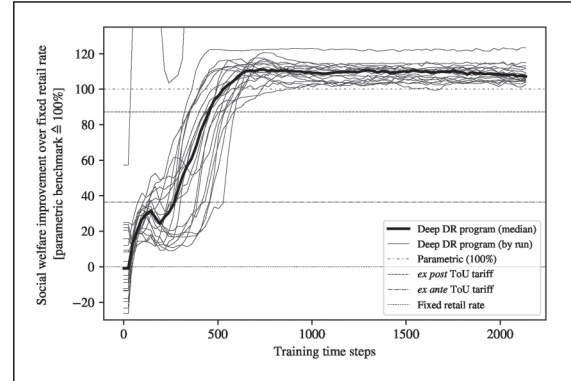


Figure 5. Convergence speed of the Deep RL algorithm. Note. Training and test data are stochastically generated and differ for all 20 runs. Improvements over the fixed retail rate are scaled to the performance of the best benchmark, that is parametric benchmark $\hat{=} 100\%$. Supplement F.3 further shows performance under identical test and training data.

19 days, on average, and after 23 days for the median run. We report both values as one parametric run performs particularly bad. Learning saturates after about 575 training time steps (24 days), that is, the program reaches 95% of its final performance. Analyzing individual runs (thin lines in the figure), we find that the best exploration run provides results better than the parametric benchmark after a learning period corresponding to 48 hours (2 days), the second best after 384 hours (16 days), and the worst one after 864 hours (36 days). This additionally demonstrates the robustness of our approach for individual runs. In Supplement F.4, we further provide details on the computational resources used as well as the necessary training time.

6.3 Impact of the Load Portfolio

We furthermore evaluate the Deep DR program for different load compositions, as documented in Table 1. The first four columns list different compositions of load types. Each compartment is dedicated to one load type (indicated in bold) which we systematically vary. Note that the shares of elastic, interruptible, and noninterruptible loads add up to 100% of demand under a fixed retail rate, while storage is measured as a share of peak demand. The second half of the columns documents the social welfare improvement in [kUSD], as achieved by the Deep DR program, and compares it to the performance of the parametric benchmark.

We find that the Deep DR program is able to realize consistent improvements across load portfolios, as compared to a fixed retail rate. For instance, for a 0% share of elastic load types in the portfolio, our program achieves an improvement of 389.73 kUSD, as compared to a situation with no DR. If the load portfolio is entirely composed of elastic loads (100%) as well as the default share of storage, an improvement of 439.29 kUSD can be achieved. Comparing our results

Table 1. Social welfare improvement in August 2020 for varying load compositions.

Load composition [%] ^a				Social welfare improvement [kUSD] ^b			
Elastic loads [%]	Interruptible loads [%]	Noninterruptible loads [%]	Storage [%]	Deep RL-based	parametric	Deep RL-based/ parametric	p value
0.00	50.00	50.00	50.0	389.73	234.24	1.66	.000
33.33	33.33	33.33	50.0	382.35	339.85	1.13	.003
66.67	16.67	16.67	50.0	407.64	387.31	1.05	.000
100.00	0.00	0.00	50.0	439.29	424.82	1.03	.000
50.00	0.00	50.00	50.0	396.53	372.27	1.07	.000
33.33	33.33	33.33	50.0	382.35	339.85	1.13	.003
16.67	66.67	16.67	50.0	381.01	301.00	1.27	.001
0.00	100.00	0.00	50.0	390.55	265.73	1.47	.000
50.00	50.00	0.00	50.0	416.13	346.18	1.20	.003
33.33	33.33	33.33	50.0	382.35	339.85	1.13	.003
16.67	16.67	66.67	50.0	376.47	255.98	1.47	.001
0.00	0.00	100.00	50.0	358.92	48.29	7.43	.000
33.33	33.33	33.33	0.0	214.00	174.05	1.23	.028
33.33	33.33	33.33	50.0	382.35	339.85	1.13	.003
33.33	33.33	33.33	100.0	557.35	470.11	1.19	.001

^aWe systematically change the shares of an elementary load type in terms of total load to 0.00%, 33.33%, 66.67%, and 100.00%, respectively, and distribute the residual load contribution equally to the other load groups. The storage share remains constant except for the last row block, where its share is systematically increased from 0% to 100% in steps of 50 percentage points. ^bSocial welfare improvement compared to the parametric benchmark. Results are documented for the month of August and averaged across 20 runs. The p value refers to the null hypothesis that the Deep DR approach does not outperform the parametric benchmark.

for different load types, we find that improvements are higher for load portfolios with higher shares of elastic loads and higher participation of storage. This can be explained by a larger flexibility potential of elastic loads and additional storage. For interruptible loads, the performance stays relatively constant. Finally, the deterioration for higher shares of noninterruptible loads could be caused by the increasing relevance of discontinuities in the response of noninterruptible loads. Such discontinuities could negatively impact the learning process.

We further compare the Deep DR program and the parametric benchmark. We find that the advantage of the Deep DR program over the parametric benchmark decreases for higher shares of elastic loads and increases for load portfolios where the other load types dominate. For instance, the advantage over the parametric benchmark is 66% if the load portfolio contains no elastic loads, but shrinks to 3% if it consists of elastic loads and storage, only. The advantage further generally increases for higher shares of storage. If no storage is present, the difference in performance of the Deep DR program and the parametric benchmark is insignificant, but the advantage of the Deep DR program increases to 19% when storage reaches 100% of peak load. We explain this finding as follows: the demand model used for the parametric benchmark works best for elastic loads but cannot represent the discontinuities and nonlinearities exhibited by the other load types.

6.4 Impact of the Notification Interval

We further investigate the impact of the notification interval on social welfare. For a notification interval of zero, this corresponds to a real-time price (RTP). For a notification interval larger than zero, load operators are notified in advance and can plan their dispatch in a cost-minimal way.

In practice, longer notification intervals are associated with deteriorating forecasts with regard to future wholesale market prices. We therefore investigate the optimal choice of the notification interval when increasing notification intervals are associated with deteriorating forecasting quality, modeled by an increasing forecasting error variance σ_{err}^2 . We analyze notification intervals of {0 (RTP); 1; ... ; 6} hours.

Figure 6(a) illustrates social welfare changes for different relationships of notification interval n and forecasting error variance σ_{err}^2 , as described in the legend. Regardless of the specific relationship, we see that welfare improves for small increases in the notification interval in comparison to a notification in real-time. However, if forecasting quality deteriorates over time, we observe that, for longer notification intervals, welfare again deteriorates and may even perform worse than notification in real-time. For a doubling of the forecasting error variance per hour ahead of dispatch (2^{n-1} and 2^n), the optimal notification interval is three and four hours, respectively. For a linear increase in the forecasting error variance ($2 \cdot n$), a notification interval of at least five hours is welfare-maximizing. In summary, this analysis demonstrates that the DR program

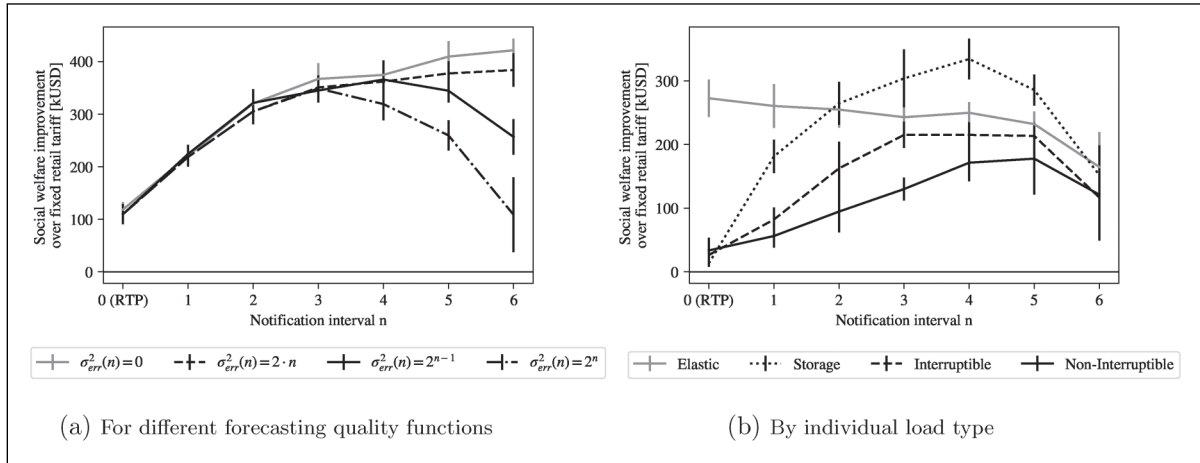


Figure 6. Welfare changes by notification interval. Note. Results for the month of August. Bars represent the 95% confidence interval over 20 runs. (a) Forecasting error variance σ_{err}^2 is 0 for $n = 0$, across all functional forms. (b) Forecasting error variance σ_{err}^2 corresponds to functional form 2^{n-1} , except for $n = 0$ where forecasting error variance is 0.

operator faces a trade-off between having access to a large flexibility potential by notifying load operators early (long notification intervals) and the availability of accurate information regarding upcoming wholesale market prices (short notification intervals). As a result, the most cost-effective notification interval is generally shorter the faster the forecasting quality deteriorates. In Supplement F.8, we further show that the relationship between the notification interval and social welfare generally persists under different customer expectations with regard to upcoming prices.

The relation between social welfare and the notification interval is more heterogeneous when differentiated along load types. Figure 6(b) shows that, if the load composition is entirely composed of elastic loads, short notification intervals lead to the highest social welfare improvements. The reason is that elastic loads do not face time-interdependent constraints. For the other load types, in contrast, a notification in real-time produces sub-optimal welfare changes. For storage, optimal welfare improvements can be reached for a notification interval of four hours. This corresponds to the time needed to fully precharge and prepare for increased DR prices. For any further increases of the notification interval, the forecasting quality deterioration dominates any possible further improvements. Finally, interruptible and noninterruptible loads are the least flexible load types. Compared to elastic loads and storage, they generally show low levels of social welfare contributions. For interruptible loads, a notification interval of three to five hours is optimal under the given specifications of the system; for noninterruptible loads, a notification interval of four to five hours produces the largest improvement.

In summary, the analyses reveal that *ex ante* information provided by the DR program operator through early notification of load operators is important. The optimal notification interval depends on the forecasting quality and load composition, both of which can vary considerably between real-world

local electricity systems. Therefore, the notification interval should be tailored to the characteristics of the system of interest. We additionally include our results in comparison to the parametric benchmark in Supplement F.9. Our approach generally outperforms the parametric benchmark, except for very long notification intervals and elastic loads, as already reported in Table 1.

6.5 Characteristics of the DR Pricing Policy

We finally leverage our numerical study to provide insights on the pricing policy of the DR program operator. Such insights can also be considered by regulators when investigating whether a DR program is effective and presumably social welfare maximizing.

First, we aim to understand how the state of the system influences price setting. We hypothesize that the longer a DR event is lasting, the more flexibility of the loads is exploited—so that higher price incentives are needed to achieve a response of the loads. Section 7 presents the average price setting (vertical axis) as a function of each hour’s position in a consecutive DR event sequence—a set of DR hours that occur back-to-back without interruption. Specifically, DR prices charged in the first hour of a consecutive event are labeled as “1st,” the second hour as “2nd,” and so on. The figure clearly demonstrates the hypothesized pattern: as the DR event window spans more consecutive hours, prices tend to rise. For instance, comparing the fitted lines of the first hour of the event sequence with the fifth hour shows that prices are generally higher the longer the event persists.

In addition, we investigate how this behavior varies with the event probability (horizontal axis), defined here as the percentage of hours within a specific hour of the day that experience a DR event. We hypothesize that hours with higher event probabilities are more likely to coincide with higher wholesale market prices, and therefore also see higher DR prices. Indeed,

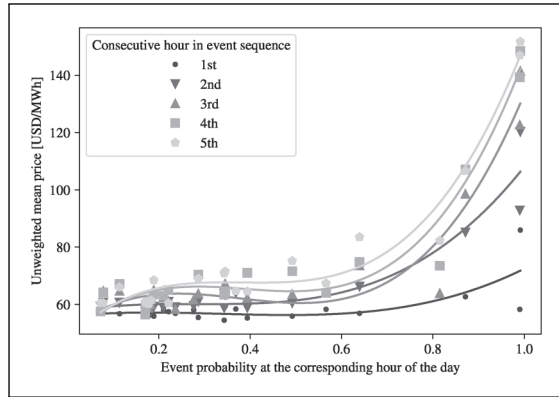


Figure 7. DR prices depending on the hour into an event sequence. Note. Results for the month of August. Event probability in a given hour is computed as the share of events in a given hour of the day. The fitted functions correspond to polynomials of degree four.

comparing price levels between low event probability hours (left end of the x -axis) and high event probability hours (right end) shows that prices increase for hours with higher event probabilities. Finally, we observe that the gap in DR prices between the first hour and, for example, the fifth hour of a consecutive event sequence becomes more pronounced as event probabilities increase.

We pursue several additional analyses on how the price setting depends on the notification interval and the load portfolio in Supplement F.10. First, we find that for very short notification intervals, DR prices amplify wholesale market price fluctuations to overcome load operators' intertemporal constraints and elicit a sufficiently strong response. For instance, when notification is in real time, DR prices increase wholesale market price changes by a factor of about 1.15 (see Figure F11). Second, we demonstrate that mean DR prices decrease as notification intervals lengthen, primarily because of increased load flexibility and the potential to shift consumption to lower-priced hours (see Figure F12). Third, we observe that the composition of the load portfolio impacts price fluctuations (see Table F4). For example, higher shares of elastic loads and storage lead to reduced DR price fluctuations.

Finally, we discuss how these insights inform DR price setting. First, DR prices should positively correlate with wholesale market prices. If the DR price does not change or converges to the maximum of the set of possible DR prices (as defined in Equation (2)), the DR program operator might deviate from social welfare maximization for the sake of maximizing private profits. Second, DR prices tend to rise with longer event durations, reflecting the need to realign intertemporal dispatch with the most valuable periods for load reductions. Third, longer notification intervals not only boost social welfare but also lower prices for customers and increase consumer surplus—provided forecasting remains sufficiently accurate. Policy makers can thus support welfare redistribution to customers by supporting longer notification intervals.

Lastly, optimal DR prices should adapt over time as the load portfolio evolves. For instance, as storage-type loads (e.g., batteries, home storage) proliferate, both DR prices and their co-movement with wholesale market prices should decrease.

7 Discussion and Conclusions

In the following, we discuss our contributions, limitations and potential challenges in real-world implementation, as well as future research directions to address these challenges.

7.1 Theoretical Contributions

First, we contribute to research by demonstrating how DR programs can be designed for groups of consumers in local electricity systems, for which constraints are unknown and individual responses are not observable to the DR program operator. Despite this limited information, we find that the Deep RL-based DR program can be effectively applied under a variety of system conditions and customers. The performance of the Deep DR program is particularly advantageous over previous parametric approaches for load portfolios with discontinuities and for systems that are subject to high wholesale market price fluctuations, as caused by intermittent renewable energies (McKinsey, 2024). Under high fluctuations, the Deep DR program can identify prices that are better aligned with the actual costs of electricity and outperforms the benchmarks. Similarly, the Deep DR program is particularly valuable if the DR program operator has access to price forecasts of high quality and can, therefore, pass on cost-reflective price signals to customers early on, using long notification intervals.

Second, the length of the notification interval for future prices is another important, but often overlooked aspect of DR program design. So far, empirical evidence on the dependence of the load response on the notification interval mostly originates from surveys (e.g., Buber et al., 2013; Fridgen et al., 2018; Taylor and Schwarz, 2000). DR program design commonly assumes that DR events are either announced day-ahead (e.g., Valogianni et al., 2020; Webb et al., 2016), in real-time (Khezeli et al., 2017), or it does not explicitly model DR events (e.g., Agrawal and Yücel, 2022; Aïd et al., 2022). We find that effective notification intervals are the result of a trade-off between time-interdependent constraints of load operators and deteriorating forecasting accuracies. Specifically, if forecasting accuracy is limited, the DR program operator cannot correct already announced DR prices when system conditions become clearer. In this case, load operators react to prices that do not reflect actual wholesale market costs and social welfare decreases. For research and practice, it is therefore relevant to assess the DR potential as a function of information availability. Many studies of load flexibility potential consider complete information (e.g., Gerke et al., 2020; Gils, 2016) and might, therefore, overestimate the ability of loads to adjust to DR events.

Finally, to design the presented DR program, we solve a novel, complex pricing problem under unknown, time-interdependent, and discontinuous demand. Within the theoretical pricing literature, current contributions consider time-interdependencies and discontinuities separately, but not in combination. Within the literature on DR program design, the majority of approaches assume a parametric form of aggregate demand or focus on specific appliances. Our results suggest that such models might impose unnecessary restrictions. Except for extreme scenarios with strong discontinuities (that result from high shares of noninterruptible loads or storages), our Deep RL-based approach is effective to solve such challenging pricing problems. It is therefore worthwhile to explore also for similar settings in other domains.

7.2 Practical Contributions and Potential Challenges Towards Real-World Implementation

DR programs are already highly relevant for today's operations of distribution systems. The design of these local DR programs is, however, based on general experience and insights from pilots rather than on a systematic methodological approach. In our numerical experiments, Deep RL has shown to identify effective prices in short learning periods from aggregate demand only with minimum requirements for communication hardware, positioning it as a promising candidate to overcome the shortage of metering infrastructure in practice.

For DR program operators, a direct requirement for an implementation in practice is to gather insights about the predictability of prices at their respective wholesale market, e.g., by evaluating external forecasting services (such as AleaSoft, 2024; DNV, 2024). The analyses of the effects of forecasting accuracies (see Supplement F.5.2) can provide them with guidance on whether or not the forecasting quality is sufficient for a successful practical implementation of the DR program.

However, some challenges remain to be solved for successful real-world implementation. First of all, guardrails need to be provided to avoid undesired outcomes. In rather rare cases, a DR policy may not converge during training or produce inferior results. This can be noticed through unusual periods of constant prices. In this case, human experts may be included into the loop to decide on reinitializing training. Moreover, exploration and pricing policies could be informed by experts and their domain knowledge (c.f. Mark et al., 2022). For instance, selected prices could be restricted to move within a certain band around the wholesale market price.

7.3 Future Research Directions

Based on our results, we see several promising directions for future research. First, the DR program could be extended to more complex electricity markets. While many markets are still regulated and subject to monopolistic retailers, pricing algorithms could be adjusted to consider competitive DR markets. Similarly, customers may be clustered into different tariff

groups. For instance, flexible consumers could be assigned to a Deep DR program, while less flexible customers could be subscribed to ToU rates, benefiting from their predictability. Another research direction aims towards accelerating the learning process, which would further increase the practical applicability of Deep DR programs. Here, strategies such as curriculum or transfer learning approaches may be explored. For instance, the pricing policy could be pretrained on synthetic and historical data (c.f. Klink et al., 2020). In that case, the DR program operator could use an approximate synthetic model of the system and later adjust the Deep RL agents to the characteristics of the real-world system.

Finally, while the presented Deep DR program solves a relevant problem in the domain of DR systems, it is also an example of how general advances in machine learning-based algorithms can help to meet the complex operational requirements of future local electricity systems. We are excited to see how this promising area develops in the near future.

Acknowledgments

The authors thank the participants of the 2024 ISDAEB Workshop, 2022 MSOM Conference, 2022 OR Conference, 2019 WITS, 2019 WCBA, 2018 Meeting of the POMS, and 2018 USAEE Conference. Marie-Louise Arlt acknowledges the support of the German Academic Scholarship Foundation.



Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

ORCID iDs

Marie-Louise Arlt  <https://orcid.org/0000-0002-2887-1794>
 Gunther Gust  <https://orcid.org/0000-0001-6671-1380>

Supplemental Material

Supplemental materials for this article are available online (doi: 10.1177/10591478251363654).

Note

1. A recent industry study estimates that future wholesale market price fluctuations will increase drastically due to renewable energies (McKinsey, 2024).

References

- Adelman D and Uçkun C (2019) Crosscutting areas—dynamic electricity pricing to smart homes. *Operations Research* 67(6): 1520–1542.
- Agrawal VV and Yücel S (2022) Design of electricity demand-response programs. *Management Science* 68(10): 7441–7456.
- Aïd R, Possamai D and Touzi N (2022) Optimal electricity demand response contracting with responsiveness incentives. *Mathematics of Operations Research* 47(3): 2112–2137.

- Al-Gwaiz M, Chao X and Wu OQ (2017) Understanding how generation flexibility and renewable energy affect power market competition. *Manufacturing & Service Operations Management* 19(1): 114–131.
- AleaSoft (2024) Energy price forecasting. <https://aleasoft.com/services/energy-price-forecasting/>.
- Anees A, Dillon T, Wallis S, et al. (2021) Optimization of day-ahead and real-time prices for smart home community. *International Journal of Electrical Power & Energy Systems* 124: 106403.
- Bahrami S, Chen YC and Wong VWS (2021) Deep reinforcement learning for demand response in distribution networks. *IEEE Transactions on Smart Grid* 12(2): 1496–1506.
- Barth L, Ludwig N, Mengelkamp E, et al. (2018) A comprehensive modelling framework for demand side flexibility in smart grids. *Computer Science - Research and Development* 33(1-2): 13–23.
- Bertsimas D and Vayanos P (2017) Data-driven learning in dynamic pricing using adaptive optimization. <https://drive.google.com/file/d/10MGPMGGWNNKTHHzg3tBvdSIft2bN1V5W/view>.
- Blonz JA (2016) Making the best of the second-best: Welfare consequences of time-varying electricity prices. *Energy Institute Working Paper Series (WP 275)*.
- Bohn RE, Caramanis MC and Schweppe FC (1984) Optimal pricing in electrical networks over space and time. *The RAND Journal of Economics* 15(3): 360–376.
- Borenstein S (2005) The long-run efficiency of real time electricity pricing. *The Energy Journal* 26(3): 93–116.
- Borenstein S (2007) Customer risk from real-time retail electricity pricing: Bill volatility and hedgability. *Energy Journal* 28(2): 111–130.
- Boßmann T and Eser EJ (2016) Model-based assessment of demand-response measures - A comprehensive literature review. *Renewable and Sustainable Energy Reviews* 57: 1637–1656.
- Buber T, Gruber A, Klobasa M, et al. (2013) Lastmanagement für systemdienstleistungen und zur reduktion der spitzenlast. *Vierteljahrshefte zur Wirtschaftsforschung* 82(3): 89–106.
- CAISO (2022) Oasis. oasis.caiso.com/.
- CAISO (2023a) California ISO peak load history 1998 through 2022. Technical report.
- CAISO (2023b) Western energy imbalance market. Technical report.
- California (2021) CA pub util code. §§ 2776, 2777.
- Cohen MC, Lobel I and Paes Leme R (2020) Feature-based dynamic pricing. *Management Science* 66(11): 4921–4943.
- Comello S and Reichelstein S (2017) Cost competitiveness of residential solar PV: The impact of net metering restrictions. *Renewable and Sustainable Energy Reviews* 75: 46–57.
- CPUC (2019) Electric rates. <https://www.cpuc.ca.gov/electricrates/>.
- CPUC (2023) 2021 Demand response monthly reports. Technical report.
- den Boer AV (2015) Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys in Operations Research and Management Science* 20(1): 1–18.
- den Boer AV and Keskin NB (2020) Discontinuous demand functions: Estimation and pricing. *Management Science* 66(10): 4516–4534.
- den Boer AV and Zwart B (2015) Dynamic pricing and learning with finite inventories. *Operations Research* 63(4): 965–978.
- DNV (2024) Power price forecasts for Europe. <https://www.dnv.com/services/power-price-forecasts-for-europe-107154/>.
- Dong C, Ng CT and Cheng TCE (2017) Electricity time-of-use tariff with stochastic demand. *Production and Operations Management* 26(1): 64–79.
- European Commission (2022) Flexibility for resilience. Technical report.
- Faruqui A and Sergici S (2010) Household response to dynamic pricing of electricity: A survey of 15 experiments. *Journal of Regulatory Economics* 38(2): 193–225.
- Feuerriegel S and Neumann D (2016) Integration scenarios of demand response into electricity markets: Load shifting, financial savings and policy implications. *Energy Policy* 96: 231–240.
- Fridgen G, Saumweber A, Seyfried J, et al. (2018) Decision flexibility vs. information accuracy in energy-intensive businesses. In: *26th European conference on information systems (ECIS)*.
- Gambardella C and Pahle M (2018) Time-varying electricity pricing and consumer heterogeneity: Welfare and distributional effects with variable renewable supply. *Energy Economics* 76: 257–273.
- Gerke BF, Gallo G, Smith SJ, et al. (2020) The California demand response potential study, phase 3.
- Gils HC (2014) Assessment of the theoretical demand response potential in Europe. *Energy* 67: 1–18.
- Gils HC (2016) Economic potential for future demand response in Germany - modeling approach and case study. *Applied Energy* 162: 401–415.
- Grimm V, Orlinskaya G, Schewe L, et al. (2021) Optimal design of retailer-prosumer electricity tariffs using bilevel optimization. *Omega* 102: 102327.
- Haider HT, See OH and Elmenreich W (2016) A review of residential demand response of smart grid. *Renewable and Sustainable Energy Reviews* 59: 166–178.
- Harsha P, Natarajan R and Subramanian D (2021) A prescriptive machine-learning framework to the price-setting newsvendor problem. *INFORMS Journal on Optimization* 3(3): 227–253.
- Khezeli K, Lin W and Bitar E (2017) Learning to buy (and sell) demand response. *IFAC-PapersOnLine* 50(1): 6761–6767.
- Kienscherf PA, Collins J, Joe-Wong C, et al. (2020) Time-dependent electricity pricing using variable announcement horizons. *Energy Informatics* 3(S1): 14.
- Kim BG, Zhang Y, Van Der Schaar M, et al. (2015) Dynamic pricing and energy consumption scheduling with reinforcement learning. *IEEE Transactions on smart grid* 7(5): 2187–2198.
- Klink P, D'Eramo C, Peters J, et al. (2020) Self-paced deep reinforcement learning. <http://arxiv.org/abs/2004.11812>.
- Kök AG, Shang K and Yücel A (2020) Investments in renewable and conventional energy: The role of operational flexibility. *Manufacturing & Service Operations Management* 22(5): 925–941.
- Li Y, Yu C, Shahidehpour M, et al. (2023) Deep reinforcement learning for smart grid operations: Algorithms, applications, and prospects. *Proceedings of the IEEE* 111(9): 1055–1096.
- Liang Z, Bian D, Zhang X, et al. (2019) Optimal energy management for commercial buildings considering comprehensive comfort levels in a retail electricity market. *Applied Energy* 236: 916–926.
- Lillicrap TP, Hunt JJ, Pritzel A, et al. (2015) Continuous control with deep reinforcement learning. <http://arxiv.org/abs/1509.02971>.
- Liu Y, Zhang D and Gooi HB (2021) Data-driven decision-making strategies for electricity retailers: A deep reinforcement learning approach. *CSEE Journal of Power and Energy Systems* 7(2): 358–367.

- Lu R, Hong SH and Zhang X (2018) A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach. *Applied Energy* 220: 220–230.
- Mark M, Chehraz N, Liu H, et al. (2022) Optimal recovery of unsecured debt via interpretable reinforcement learning. *Machine Learning with Applications* 8: 100280.
- Märkle-Huß J, Feuerriegel S and Neumann D (2018) Large-scale demand response and its implications for spot prices, load and policies: Insights from the german-austrian electricity market. *Applied Energy* 210: 1290–1298.
- McKinsey (2024) Global energy perspective 2024. Technical report.
- Mieth R and Dvorkin Y (2020) Online learning for network constrained demand response pricing in distribution systems. *IEEE Transactions on Smart Grid* 11(3): 2563–2575.
- Moazeni S, Defourny B and Wilczak MJ (2020) Sequential learning in designing marketing campaigns for market entry. *Management Science* 66(9): 4226–4245.
- Nikzad M, Mozafari B, Bashirvand M, et al. (2012) Designing time-of-use program based on stochastic security constrained unit commitment considering reliability index. *Energy* 41(1): 541–548.
- ORNL (2023) Homeland infrastructure foundation level database (HIFLD). Technical report.
- Ponse K, Kleuker F, Fejér M, et al. (2024) Reinforcement learning for sustainable energy: A survey. *arXiv preprint arXiv:2407.18597*.
- pwc (2022) Smart-Meter-Roll-out—Standortbestimmung der grundzuständigen Messstellenbetreiber. (Transl.: Smart meter roll-out - Resume of responsible metering infrastructure operators).
- Qian T, Liang Z, Shao C, et al. (2024) Offline DRL for price-based demand response: Learning from suboptimal data and beyond. *IEEE Transactions on Smart Grid* 15(5): 4618–4635.
- Sarfaraizi S, Mohammadi S, Khastieva D, et al. (2022) Optimal real-time pricing for aggregating prosumagers and electric vehicles in energy communities under uncertainty. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4131119.
- Shoreh MH, Siano P, Shafie-khah M, et al. (2016) A survey of industrial applications of demand response. *Electric Power Systems Research* 141: 31–49.
- Sunar N and Birge JR (2019) Strategic commitment to a production schedule with uncertain supply and demand: Renewable energy in day-ahead electricity markets. *Management Science* 65: 714–734.
- Sunar N and Swaminathan JM (2021) Net-metered distributed renewable energy: A peril for utilities? *Management Science* 67(11): 6716–6733.
- Sutton RS and Barto AG (2020) *Reinforcement Learning: An Introduction, 2nd Edition*. Cambridge: The MIT Press.
- Tang C, Qin Y, Wu F, et al. (2024) Dynamic demand-aware power grid intelligent pricing algorithm based on deep reinforcement learning. *IEEE Access* 12: 75809–75817.
- Taylor TN and Schwarz PM (2000) Advance notice of real-time electricity prices. *Atlantic Economic Journal* 28(4): 478–488.
- US Department of Energy (2006) Benefits of demand response in electricity markets and recommendations for achieving them. Technical report.
- US Department of Energy (2022) Demand response in industrial facilities: Peak electric demand. Technical report.
- Vallés M, Bello A, Reneses J, et al. (2018) Probabilistic characterization of electricity consumer responsiveness to economic incentives. *Applied Energy* 216: 296–310.
- Valogianni K, Ketter W, Collins J, et al. (2020) Sustainable electric vehicle charging using adaptive pricing. *Production and Operations Management* 29(6): 1550–1572.
- Vardakas JS, Zorba N and Verikoukis CV (2015) A survey on demand response programs in smart grids: Pricing methods and optimization algorithms. *IEEE Communications Surveys & Tutorials* 17(1): 152–178.
- Vázquez-Canteli JR and Nagy Z (2019) Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Applied Energy* 235: 1072–1089.
- Wang X, El-Farra NH and Palazoglu A (2017) Optimal scheduling of demand responsive industrial production with hybrid renewable energy systems. *Renewable Energy* 100(2017): 53–64.
- Webb EM, Wu OQ and Cattani KD (2016) Coordinating energy efficiency and incentive-based demand response. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2843798.
- Wu X, Yang M, Liang L, et al. (2025) Redesign of virtual power plant-led electricity demand response program. *Production and Operations Management* 34(7): 1817–1835.
- Xu H, Hu Q, Wu Q, et al. (2024) Deep multi-task multi-agent reinforcement learning based joint bidding and pricing strategy of price-maker load serving entity. *IEEE Transactions on Power Systems* 40(1): 505–517.
- Xu H, Sun H, Nikovski D, et al. (2019) Deep reinforcement learning for joint bidding and pricing of load serving entity. *IEEE Transactions on Smart Grid* 10(6): 6366–6375.
- Xu H, Wen J, Hu Q, et al. (2022a) Energy procurement and retail pricing for electricity retailers via deep reinforcement learning with long short-term memory. *CSEE Journal of Power and Energy Systems* 8(5): 1338–1351.
- Xu H, Wu Q, Wen J, et al. (2022b) Joint bidding and pricing for electricity retailers based on multi-task deep reinforcement learning. *International Journal of Electrical Power & Energy Systems* 138: 107897.
- Yan X, Ozturk Y, Hu Z, et al. (2018) A review on price-driven residential demand response. *Renewable and Sustainable Energy Reviews* 96: 411–419.
- Yousefi S, Moghaddam MP and Majd VJ (2011) Optimal real time pricing in an agent-based retail market using a comprehensive demand response model. *Energy* 36(9): 5716–5727.
- Zhang Z, Zhang D and Qiu RC (2020) Deep reinforcement learning for power system applications: An overview. *CSEE Journal of Power and Energy Systems* 6(1): 213–225.
- Zhou E and Mai T (2021) Electrification futures study: Operational analysis of U.S. power systems with increased electrification and demand-side flexibility. Technical report.
- Zhou Y, Scheller-Wolf A, Secomandi N, et al. (2016) Electricity trading and negative prices: Storage vs. disposal. *Management Science* 62(3): 880–898.
- Zhou YH, Scheller-Wolf A, Secomandi N, et al. (2019) Managing wind-based electricity generation in the presence of storage and transmission capacity. *Production and Operations Management* 28(4): 970–989.

How to cite this article

Arlt M-L, Gust G and Neumann D (2026) A Deep Demand Response Program for Local Electricity Systems. *Production and Operations Management* 35(3): 1020–1037.