Price-Based Demand Response Program Design in Constrained Electricity Systems

Marie-Louise Arlt* Gunther Gust † Dirk Neumann‡

June 30, 2020

Abstract

Renewable energies require new approaches to the operation of power systems. In this work, we present a novel Demand Response (DR) program to mitigate grid congestion. The program stands out by adjusting electricity prices online and solely based on observed aggregate electricity consumption. It thereby avoids privacy concerns and expensive information system infrastructure investments. Additionally, it can cope with non-elastic and time-interdependent demand—which is typical for storage devices and industrial processes. Beyond the DR program, this paper contributes to research on flexible load modeling. Specifically, we model four flexible load types and provide a unified framework of load modeling under uncertainty and with time-interdependencies. Numerical experiments show that our DR programachieves considerable and stable cost savings of 40% to 60% in comparison to conventional DR programs. Moreover, the solution approach based on Deep Reinforcement Learning reaches these savings after a short learning period corresponding to only 25 simulation days. Regarding load flexibility, we find that the responses of the four load types strongly differ depending on prices and the lag between the announcement and the application of price changes ('notification interval') which is an important design parameter. Our results imply that system operators should flexibilize their DR programs with the help of learning algorithms and tailor their program to the local load composition, congestion frequency, and forecasting quality. For researchers, our solution approach and our unified framework can help to solve pricing problems with unknown and heterogeneous demand, which occur on online platforms and in other environments.

Keywords: *Dynamic pricing, Demand response, Electricity, Demand flexibility, Time-interdependent demand*

*Albert Ludwig University of Freiburg, Stanford University, SLAC National Accelerator Laboratory
†Albert Ludwig University of Freiburg
‡Albert Ludwig University of Freiburg

1 Introduction

Renewable energy systems are characterized by increasingly distributed generation and dynamically changing power flows. This challenges conventional approaches to system operations and requires a leveraging of the flexible load potential in order to meet the energy balance and avoid network congestion. Demand Response (DR) is an established concept for addressing several important operational and managerial problems in electricity systems (Bollinger and Hartmann, 2020; Parker et al., 2019), such as cost-efficient energy procurement (Adelman and Uçkun, 2019) and operation under grid constraints (Mieth and Dvorkin, 2020b). Thereby, DR programs contribute to the efficient long-term management of energy supply (Kök et al., 2018). Among possible concepts of DR, price-based DR programs incentivize electricity customers to deviate from their original consumption patterns through time-varying electricity prices (US Department of Energy, 2006).

Despite their wide-spread deployment, existing price-based DR programs do not sufficiently consider the challenges of future electricity systems. First, the majority of practically implemented programs like time-of-use rates are not sufficiently adaptive to dynamic renewable infeed and potential grid congestion. Second, more flexible programs as suggested by academic contributions often require detailed information about loads which, in practice, is not available for privacy considerations and other reasons. Third, existing DR programs do not sufficiently consider the dynamic nature of the decision problem of load operators. Loads are generally subject to time-interdependent constraints and therefore state-dependent (e.g. with respect to the state-of-charge of a battery) which limits their ability to respond. Particularly in local electricity systems with only a few loads, state dependence must not be neglected because the behavior of individual loads can dominate overall DR. The majority of publications model neither the arrival of information nor the state dependence. Fourth, existing approaches determine DR prices based on information about the aggregate system level. Information, such as wholesale market prices (e.g. Chen et al., 2013; Celebi and Fuller, 2007), is used to estimate price sensitivities; even though they hardly reflect the local demand elasticities nor do they reflect capacity constraints in the local electricity system. Using this information can produce misleading recommendations for the parametrization of a DR program.

To address these shortcomings, we present a novel DR program for constrained local electricity systems. Our work contributes to the research on DR programs in three ways. First, we propose a DR program which allows for online DR price setting by the system operator (SO). This allows for the consideration of dynamic system challenges in modern electricity systems, such as temporarily changing congestion. Second, our solution method based on Deep 1RL requires no information about demand except for realized electricity consumption and can cope with non-elastic and time-interdependent demand. This considerably relaxes assumptions of previous work. Third, we explicitly model the dynamic information structure of DR programs. Specifically, we include a 'notification interval' in our model, i.e. the duration between the announcement and the application of a price change during a congestion event.

Furthermore, our work provides three contributions regarding the modeling of flexible loads for DR programs and beyond. First, we adopt a bottom-up approach based on four elementary load types – storage, interruptible and non-interruptible loads, as well as elastic loads – to model system load. This allows us to consider more complex load behavior. Second, we investigate the impact of state dependence on the amount of DR. Third, we model the decision process of load operators using the language of optimal control under uncertainty. We thereby provide a unified modeling framework to represent the four elementary load types. Due to the generic nature of the framework, it can serve researchers to model flexible loads in applications beyond DR and the electricity sector.

Our research has promising managerial implications. First, SOs should work towards flexibilizing their conventional fixed DR price programs. Our variable DR price program achieves considerable cost savings in comparison to existing fixed DR price programs. Second, SOs should begin to test Deep RL as a tool for congestion management. Our solution method based on Deep RL is able to provide stable cost savings within a short training time, without requiring information on the underlying load structure. Therefore, it resolves privacy concerns and avoids expensive information technology investments for communicating with load operators. Third, as an immediate implication, we provide several guidelines for the direct improvement of existing DR programs, e.g. on how to tailor them to the local load composition, forecasting qualities, and congestion frequencies.

Moreover, our work is also relevant for policymakers and regulators. Our variable DR price program has several advantages over competing concepts, such as distribution locational marginal prices. Among others, it limits the price risk for consumers, avoids undesired price differentiation between customers, and does not require the establishment of a local market platform. Furthermore, while it is challenging to regulate prices under the temporally and locally variable requirements of electricity systems, regulators can alternatively approve the deterministic price-setting neural network underlying our solution approach.

We proceed as follows. In Section 2, we summarize the related literature on DR programs and flexible load modeling and describe the relevant research gaps. In Section 3, we present our model, including the cost minimization problems of both the SO and the flexible load operators. We thereby show that the resulting aggregate load function of the load operators exhibits time-interdependencies, discontinuities, and inelasticities which make the DR program very challenging to optimize. In Section 4, the solution method based on Deep RL is explained. We evaluate the results in a numerical experiment in Section 5. We conclude our work by discussing the implications for management, policy, and research in Section 6.

2 Related Work

2.1 Demand Response Programs

Among the large body of research on DR programs (as summarized in e.g. Deng et al., 2015; Albadi and El-Saadany, 2008; Boßmann and Eser, 2016), we restrict our review to price-based DR programs as defined by the US Department of Energy (2006). Also, we focus on contributions which include the problem of determining prices, in contrast to the large amount of research that analyzes the performance of a DR program with given prices (such as Faruqui and Sergici (2010) for residential loads and Shoreh et al. (2016) for industrial loads).

Existing price-based DR programs use mostly fixed prices. Examples include time-of-use rates and critical peak pricing programs (US Department of Energy, 2006). In both time-of-use rates and critical peak pricing programs, prices are fixed *ex ante* and constant during peak time periods. Time-of-use rate programs additionally fix the timing of peak periods. Both programs provide the SO with very limited flexibility to address load peaks varying in time and size. An alternative would be the use of distribution locational marginal prices which reflect the local value of electricity in real-time (e.g. Sotkiewicz and Vignolo, 2006; Liu et al., 2018; Mieth and Dvorkin, 2020a). Distribution locational marginal prices might, however, expose consumers to unwanted price risks and price differentiation among customers (e.g. US Department of Energy, 2006; Borenstein, 2007; Zarabie et al., 2019) which can be problematic from a policy perspective. As an additional downside, distribution locational marginal prices require an extensive information and market system, e.g. Mengelkamp et al. (2018). Alternatively, more flexible DR programs can combine the low price risk of conventional DR programs and the responsiveness of distribution locational marginal prices to dynamic constraints.

These more flexible DR programs differentiate prices according to various factors, such as the hour of the day, and may even determine prices online. Examples closely related to our work include Adelman and Uçkun (2019) who propose a dynamic pricing approach for heating, ventilation, and air conditioning systems; Valogianni et al. (2020) who suggest adaptive pricing for electric vehicles; and Mieth and Dvorkin (2020b) who propose a time-specific DR program optimization for distribution grids. Similar to these approaches, we propose a DR program that determines prices online and thereby combines the flexibility of distribution locational marginal prices with the advantages of conventional DR programs. However, our DR program stands out from the literature, including the aforementioned approaches, regarding several key aspects: most importantly, the absence of information to the SO, the consideration of several load types, the state dependence of load, and the consideration of a notification interval. We explain these characteristics below.

In order to determine variable prices, the literature on DR programs makes different assumptions regarding the knowledge of the SO about demand: part of the literature assumes full information regarding the price elasticities (e.g. Kirschen et al., 2000; Nikzad et al., 2012; Doostizadeh and Ghasemi, 2012) or the number and types of loads (e.g. Chen et al., 2012). Full knowledge of the load response, however, is rarely given – due to a lack of communication infrastructure, privacy considerations, or other reasons. Therefore, another part of the literature assumes that the price setting entity has no or only limited information about demand functions. In this kind of setting, learning algorithms can be a promising solution approach, as argued by den Boer (2015) for markets beyond electricity. Recent examples are, for instance, provided by Moazeni et al. (2019) and Cohen et al. (2020) in marketing, and Papanastasiou (2020) for inventory management. As for electricity applications, Khezeli et al. (2017) and Mieth and Dvorkin (2020b) propose least squares and quantile estimation for optimizing prices in a DR program. Bompard and Han (2013) consider a market-based control for customers with evolving comfort and price sensitivities. Valogianni et al. (2020) suggest an adaptive learning algorithm based on a quadratic utility function. Furthermore, Lu et al. (2018) and Yousefi et al. (2011) propose Q learning approaches. While the SO in Lu et al. (2018) is able to measure demand of individual loads, the agent in Yousefi et al. (2011) only observes aggregate DR. Generally, all these contributions focus on time-independent and differentiable demand functions. Real systems, however, differ in two important aspects: first, loads show effects of inter-temporal substitution and, second, their response is not necessarily proportional to price changes. In our work, we assume that no other information about the system but aggregate load is available to the SO and we do not require continuous and differentiable demand functions. For this purpose, we later present a solution based on Deep RL that is able to identify effective DR prices under these challenging circumstances.

In addition to the absence of information on demand, the price setting problem is inherently dynamic. Some information only arrives intraday (e.g. on upcoming load and generation) and, therefore, system-relevant information is only completely available at real-time. However, providing information about a DR event just before dispatch generally decreases the ability of flexible load to respond, as Buber et al. (2013) showed in an industry survey. Accordingly, Fridgen et al. (2018) find decreasing energy procurement cost for a non-interruptible load under longer planning horizons. Taylor and Schwarz (2000) have made a first attempt in modeling the relationship between the ability of loads to respond and the uncertainty of system operations. The authors represent loads as fully elastic and compare costs under a notification in advance and in real-time. In their analysis, the authors assume load elasticities and their increase with a longer notification interval as given. In our work, we extend their analysis by determining load response endogenously as a function of the notification interval and time-interdependent constraints, instead of a given elasticity. Furthermore, we endogenously determine effective DR prices and incorporate imperfect forecasting quality. This allows us to not only quantify the effect of the dynamic arrival of information but also to find strategies to effectively address the trade-off between uncertainty about future events and the ability of loads to respond. To the best of our knowledge, our work is the first approach that models the dynamic arrival of information explicitly and analyzes the effect of the length of the notification interval on DR program performance.

2.2 Flexible Load Modeling in DR programs

In order to evaluate DR programs, a model of flexible load is required. The majority of research on DR program design is conducted at an aggregate level, e.g. at the level of transmission networks. These approaches (e.g. Dong et al., 2017; Zhang, 2014; Chen et al., 2013; Celebi and Fuller, 2007) derive prices from the wholesale market or from aggregate demand elasticity estimations. At lower levels of aggregation, such as the distribution level, individual loads can dominate aggregate load behavior. This renders the common assumptions of aggregate approaches of continuous load adjustment and constant elasticities invalid. Such aggregate analyses might lead to misleading results for local electricity systems. Therefore, modeling electricity consumption as the sum of individual behavior by bottom-up approaches is more adequate. The few examples of bottom-up approaches include Adelman and Uçkun (2019) who consider heating, ventilation, and air conditioning systems and Mieth and Dvorkin (2020b) who assume agents with continuous cost and utility functions. We follow these contributions using a bottom-up approach but consider multiple load types – namely storage, interruptible, non-interruptible, and elastic loads. We show later in the paper that these load types respond very differently to dynamic prices; in particular, they exhibit different degrees of inelasticities and discontinuities. We demonstrate that effective DR programs need to take these differences into account.

Using a bottom-up approach furthermore allows us to model the state dependence of different load types, e.g. the state-of-charge of a battery or whether a non-interruptible load has been started yet. So far, state dependence has only insufficiently been covered in the literature. Current approaches for DR parametrization model load behavior either without state dependence (e.g. Mieth and Dvorkin, 2020b) or use within- and between-hour price-elasticities (e.g. Dong et al., 2017; Nikzad et al., 2012; Aalami et al., 2010; Yousefi et al., 2011). The latter are usually not available, as their reliable estimation requires that the system has been exposed to significant price variation in the past. There are only a few examples which include state dependence in their analysis: Adelman and Uçkun (2019) introduce a global and a local state, representing weather (global) and internal temperature (local). Sioshansi (2012) focuses on electric vehicles and their state-of-charge. Wang et al. (2017) propose an incentive-compatible pricing mechanism for individual loads, using a chlor-alkali process with hydrogen storage and a fuel cell as a case study for which dispatch is restricted by the state of the hydrogen tank. These approaches, however, do not represent the diversity of loads present in real-world electricity systems, including the variety of possible state-binding constraints. In contrast, we extend these single-load analyses to the four elementary load types mentioned previously. We thereby quantitatively show for the first time the effect of different degrees of state dependence on the load response and on effective DR prices.

2.3 Flexible Load Modeling

Since we model state-dependent load behavior for four load types, our research is also relevant for the stream of literature on flexible load modeling beyond DR programs. Within this literature, contributions covering several load types include, e.g., Petersen et al. (2013), Barth et al. (2018), Xu et al. (2016), Mitra et al. (2012), and Schott et al. (2019). Petersen et al. (2013) and Xu et al. (2016) both suggest modeling frameworks for flexible loads which are operated by a central dispatcher. Mitra et al. (2012) propose a generalized model to represent flexible industrial loads for dispatch planning. Barth et al. (2018) take the perspective of load operators to represent the deterministic dispatch of different load types. We follow Barth et al. (2018) and propose objective functions which reflect the individual cost-minimizing behavior of load operators. Our approach, however, extends this framework. By relying on optimal control under uncertainty, we are able to model uncertainty and expectations about the future dynamics of the system. In the context of our DR program, our load model enables us to capture the impact of sequential information on the system and provide insights on how the different load types are differently affected. Finally, the unified modeling framework can serve researchers to model flexible demand in a variety of applications beyond DR and the electricity sector.

3 The Model

We consider an 1SO who is operating a local electricity system. Local electricity systems are composed of only a limited number of load operators such that individual dispatch decisions impact aggregate load in the system. If grid congestion is present, the SO can deviate from the base electricity price through a price-based DR program. By doing so, the SO can incentivize a change in load behavior instead of dispatching expensive reserve capacities or curtailing loads.

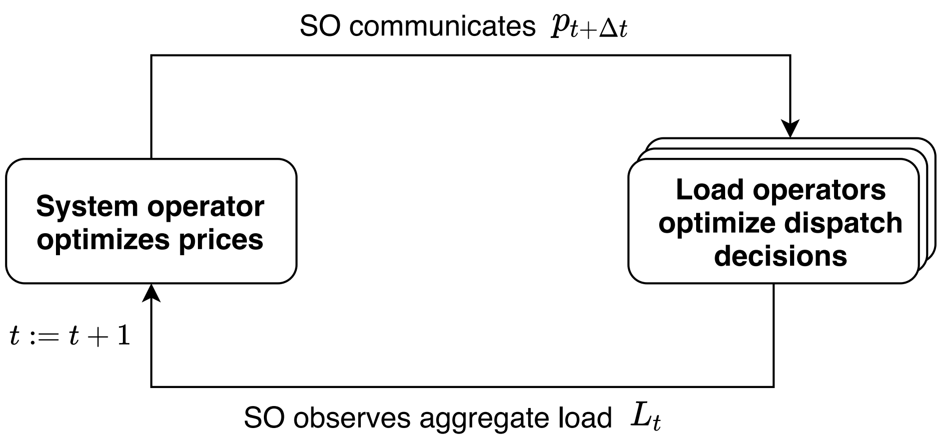


Figure 1: Agents and sequence of actions in period t of our variable DR price program

The DR program is modeled as follows. We represent the SO and the load operators as cost-minimizing agents which make sequential decisions in discrete time, as illustrated by Figure 1. If, in period t, congestion is forecasted for period $t + \Delta t$, the SO can choose to deviate from the base price $p^b_{t+\Delta t}$ and set a price $p_{t+\Delta t} \neq p^b_{t+\Delta t}$ instead. We refer to Δt as the *notification interval*. Prices for periods $\tau \in \{t, ..., t + \Delta t\}$ are known to load operators at time t; for periods $\tau > t + \Delta t$, prices are uncertain. The load operators observe $p_{t+\Delta t}$, re-optimize their load schedule, and implement their dispatch scheduled for t. At the end of period t, the SO observes the resulting aggregated load L_t in the system. The SO neither knows the number of load operators nor does he observe the dispatch of individual loads. The process then repeats for time step $t + 1$.

In the section below, we formalize the decision problems of the SO (Section 3.1) and the load operators (Section 3.2). All variables and parameters are described in Table 1.

3.1 Decision Problem of the System Operator

In each period, the SO seeks to minimize his expected average stage-wise system operation costs by the choice of the future electricity price $p_{t+\Delta t}$, as given by Equation (1),

$$\min_{p_{t+\Delta t}} \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_q \sum_{\tau=t}^{t+T-1} cong_\tau \cdot |L_\tau(\overrightarrow{p}) - L_\tau^{target}| \cdot \begin{cases} P_\tau & \text{for} \quad L_\tau(\overrightarrow{p}) - L_\tau^{target} \geq 0; \\ f_\tau & \text{for} \quad L_\tau(\overrightarrow{p}) - L_\tau^{target} < 0. \end{cases} \tag{1}$$

System operation costs apply when the system is congested, i.e. when the binary variable $cong_t$ equals 1. If this is the case, the SO can minimize system operation costs by incentivizing load operators to reduce aggregate load $L_t(\vec{p})$ to meet a target load, L_t^{target} through the choice of the DR price $p_{t+\Delta t}$. If the DR programs fails to reduce aggregate load $L_t(\vec{p})$ to below the target load L_t^{target}, the SO has to pay penalties P_t for any excess load. The penalty reflects the cost of additionally dispatched reserves or curtailments. If the SO reduces load beyond L_t^{target}, a fee of f_t applies. The fee prevents the SO – who is a monopolist in the local electricity system – from setting prices too high. This situation is associated with welfare losses for consumers as the resulting allocation deviates from the efficient equilibrium. These losses have to be internalized by means of regulation to align the SO's incentives. In Appendix 7, we show that the aggregate welfare loss of load and suppliers as a result of a constrained system can be approximated by $f_t = p_t$ as an upper bound. Likewise, the problem can also be defined for a generation-constrained system, e.g. as a result of high renewable energy generation, as shown in Appendix 8.

The optimization problem of the SO is an optimal control problem under uncertainty (Bertsekas, 2000). In each period t, the SO determines the DR price $p_{t+\Delta t}$, which minimizes his expected system operation costs, given the notification interval Δt. While doing so, the SO faces two unknowns: first, the occurrence of a congestion event in $t + \Delta t$ is subject

Symbol	Description	Range
System		
T	Discrete time horizon	$T \in \mathbb{N}$
t, τ	Time indices	$t, \tau \in 0, ..., T-1$
Δt	Notification interval for future price changes	$\Delta t \in \{\mathbb{N}_0 \| \Delta t << T\}$
$cong_t$	Presence of congestion in t (no/yes)	$cong_t \in \{0, 1\}$
p_t	Electricity price in t (decision variable of the SO)	$p_t \in \{\mathbb{R}^+_{\neq 0} \| p^{min} \leq p_t \leq p^{max}\}$
$\vec{p}$	Vector of electricity prices	$\vec{p} = \{p_0, ..., p_{T-1}\}$
p^b_t	Base electricity price in t	$p^b_t \in \mathbb{R}^+_{\neq 0}$
$L_t(\vec{p})$	Aggregate price-dependent load in t	$L_t(\vec{p}) \in \mathbb{R}$
L^{target}_t	Target load under congestion in t	$L^{target}_t < L_t(\vec{p}^b)$
P_t	Penalty for not achieving the target load in t	$P_t \in \mathbb{R}^+_0$
f_t	Fine for over-achieving the target load in t	$f_t = p_t$
q_t	Probability of correctly forecasting congestion in t	$q_t \in [0.5, 1.0]$
Loads		
k	Type of flexible load	$k \in \mathbb{K} = \{el, sto, inter, ninter\}$
N_k	Set of flexible loads of type k	$\|N_k\| \geq 0, \forall k \in \mathbb{K}$
n	Flexible load	$n \in N_k$, $\forall k \in \mathbb{K}$
$\vec{L}_n$	Load profile of a flexible load n	$\vec{L}_n \in (\mathbb{R}^+_0)^{\|\vec{L}\|}$
L^j_n	jth component of load profile of a flexible load n	$L^j_n \in \mathbb{R}^+_0, \forall j \in \{0, ..., \|\vec{L}\| - 1\}$
x_t	State variable of a flexible load	$x_t \in \{\mathbb{R}^+_0 \| x^{min} \leq x_t \leq x^{max}\}$
x^{min}, x^{max}	Minimum and maximum state	$x^{min} = 0$, $x^{max} > 0$
$L_{n,t}(x_t)$	State-dependent flexible demand of load n in t	$L_{n,t}(x_t) \in \mathbb{R}^+_0$
u_t	Decision variable for the dispatch of a flexible load in t	Load type specific
u^{crit}_t	Level of dispatch in t for which the marginal dispatch period changes	$u^{crit}_t \in (0, 1)$
t^{start}, t^{end}	Earliest start time, latest end time	$t^{start} < t^{end} < T$
Δt^{crit}	Critical number of periods for completing the execution of a flexible load	$\Delta t^{crit} \leq t^{end} - t + 1$
ρ	State coupling parameter	$\rho \in [0, 1]$
ϵ	Price sensitivity	$\epsilon \in \mathbb{R}^+_{\neq 0}$
b_t	Additional dispatch cost in t	$b_t \in \mathbb{R}^+_0$

Table 1: Variables and parameters

to an uncertainty in forecasting, where congestion forecasting quality is quantified by q_t. Second, the SO knows neither the number of load operators, their load types, nor their dispatch optimization problems, and must therefore take his decision based on the estimated load $\mathbb{E}\, L_t(\vec{p})$. After the price has been announced, the SO observes aggregate load response, which he uses to update his beliefs about future load responses. If no congestion is forecasted, the price is restricted to the base price, i.e. $p_t = p^b_t$.

We characterize the aggregate flexible load $L_t(\vec{p})$ as a sum of the contributions of four elementary flexible load types which are based on Barth et al. (2018): storage, interruptible and non-interruptible loads, as well as elastic loads. Figure 2 illustrates the four types and their ability to respond to variable prices. Storage is able to charge and shift load into other time periods. Meanwhile, interruptible and non-interruptible loads both have pre-defined 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. For an elastic load, the response depends on the price in t only and is not time-interdependent.

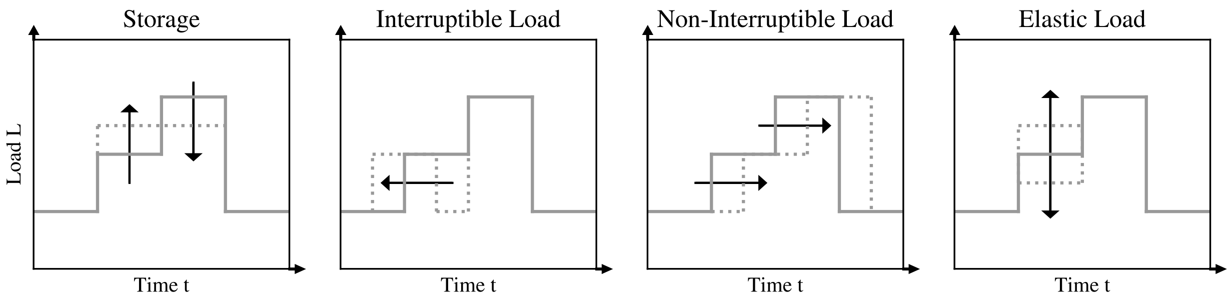


Figure 2: Flexible load types

Overall, aggregate flexible load in t is given in Equation (2). It is the sum of the cost-minimizing dispatch decision of the flexible load set N_k of each type $k \in \{sto, inter, ninter, el\}$, given their current state vectors $\vec{x}_t$ and prices $\vec{p}$,

$$L_t(\vec{x}_t, \vec{p}) = \sum_{n \in N_{sto}} L_{n,t}(x_{n,t}, \vec{p}) + \sum_{n \in N_{inter}} L_{n,t}(x_{n,t}, \vec{p}) + \sum_{n \in N_{ninter}} L_{n,t}(x_{n,t}, \vec{p}) + \sum_{n \in N_{el}} L_{n,t}(x_{n,t}, \vec{p}). \tag{2}$$

Inflexible load does not participate in the DR program and is not subject to price changes. Therefore, we do not consider it in the DR program optimization.

3.2 Decision Problem of Load Operators

Operators of storage as well as interruptible and non-interruptible loads are facing an intertemporal optimization problem when minimizing expected stage-wise electricity costs. As the SO sequentially communicates future prices, the load operators take new information on prices into account and decide upon optimal load activity in t. The objective function of a load operator with a single flexible load of type $k \in \{sto, inter, ninter\}$ reads as follows,

$$\min_{u_t} \lim_{T \to \infty} \frac{1}{T} \mathbb{E}\{ \sum_{\tau=t}^{t+T-1} (p_\tau + b_\tau) L_\tau(x_\tau, \vec{p}) u_\tau \}. \tag{3}$$

Load operation costs are driven by two components, the expected electricity price p_t and additional dispatch costs b_t, e.g. costs to shift load. Both are charged per unit of energy. The decision variable u_t describes load activity, such as charging, discharging, dispatch, or no dispatch. Because of the chosen cost structure, the load decisions are separable. The aggregate load behavior can therefore be modeled as the sum of the dispatch decisions of single load operator agents with a single flexible load each.

The state of a load is characterized by the amount of energy $x_t \in \mathbb{R}_0^+$ that has already been dispatched or that is accumulated in storage. Equation (4) describes the transition function of states and serves as an inter-temporal state constraint,

$$x_{t+1} = (1 - \rho) x_t + L_t(x_t) u_t \ , \forall t \in \{0, 1, ..., T-2\}. \tag{4}$$

The future state x_{t+1} depends on the current state x_t and the current energy consumption $L_t(x_t)u_t$. Furthermore, additional constraints apply depending on the specific flexible load type k,

$$\mathbb{C}_k, \ \forall k \in \{sto, inter, ninter\}. \tag{5}$$

which are specified in the following Sections 3.2.1 to 3.2.3.

As for the SO, we describe the scheduling problem as an optimal control problem under uncertainty to capture the dynamic decision structure of the rolling time horizon. We later determine the optimal dispatch by applying a stochastic dynamic programming approach, using Bellman's 'Principle of Optimality' by recursively solving the stages of the optimization problem.

	Storage	**Interruptible Load**	**Non-Interruptible Load**
Objective function	$\min_{u_t} \lim_{T\to\infty} \frac{1}{T}\,\mathbb{E}\{\sum_{\tau=t}^{t+T-1}(p_\tau + b_\tau)L_\tau(x_\tau)u_\tau\}$		
Transition function	$x_{t+1} = (1-\rho)x_t + L_t(x_t)u_t$		
State variable x_t	State-of-charge $x_t \in \mathbb{R}_0^+$	Cumulative energy consumption $x_t \in \mathbb{R}_0^+$	
Decision variable u_t	Charge control $u_t \in [-1, 1]$	Load activity $u_t \in \{0, 1\}$	
Load L_t	Charging rate $L_t(x_t) = L \in \mathbb{R}_0^+$	$\vec{L}$, $L_t(x_t) = L(\sum_{\tau=t^{start}}^{t-1} L_\tau u_\tau)$	
State coupling ρ	Storage losses with $0 \leq \rho^{loss} << 1$	$\rho = 0$	
Additional dispatch cost b_t	$b_t = 0$	$b_t \in \mathbb{R}_0^+$	
Additional constraints $\mathbb{C}_k$	$x_t \in [x^{min}, x^{max}]$	$x_{t^{start}} = 0$; $x_{t^{end}} = \sum_j L^j$	$x_{t^{start}} = 0$; $x_{t^{end}} = \sum_j L^j$; Non-interruptibility $u_t \leq u_{t+1} + \frac{x_t}{\sum_j L^j}$

Table 2: Unified control framework of flexible loads

In Sections 3.2.1 to 3.2.3, we apply this generic problem formulation and solution procedure to our three time-interdependent elementary load types. We thereby also analytically derive optimal dispatch policies and show that the optimal policies exhibit time-interdependencies, discontinuities, and inelasticities. This has important consequences for the aggregate load function and makes the decision problem of the SO challenging to solve. Table 2 provides an overview of the mathematical modeling of the load types in a unified framework of optimal control under uncertainty. All three load types share the previously introduced objective function as well as the transition function. Their interpretation with regard to state and decision variables, load, as well as state coupling, however, differs. The same is true for the specification of the costs for exercising flexibility as well as the constraint set. In the following section, we provide a detailed explanation of the modeling of each load type. In Section 3.2.4, we close by characterizing the dispatch behavior of an elastic load. This load type is covered separately since it does not exhibit state-dependent behavior.

3.2.1 Storage

Storage is characterized by the ability to charge electricity and discharge it at a later point in time. u_t denotes the charging decision by the storage operator, with $0 < u_t \leq 1$ for charging and $-1 \leq u_t < 0$ for discharging. L represents the maximum charging rate. x_t describes the state-of-charge of the storage at the beginning of the period t. Storing electricity results in losses relative to the total energy stored, defined by $0 < \rho^{loss} << 1$. In sum, these definitions turn the generic transition function of Equation (4) into the following storage-specific transition function,

$$x_{t+1} = (1 - \rho^{loss})x_t + Lu_t, \ \forall t \in \{0, 1, ..., T-2\}. \tag{4, sto}$$

The state-of-charge x_t is furthermore restricted by the following storage-specific constraint $\mathbb{C}_{sto}$,

$$x_t \in [x^{min}, x^{max}], \ \forall t \in \{0, 1, ..., T-1\}. \tag{5, sto}$$

It reflects the technical constraints of the storage unit, i.e. its minimum and maximum state-of-charge.

Using dynamic programming, we can derive the state-dependent charging policy given by Proposition 3.1.

Proposition 3.1 *Let $i \in \mathbb{N}^+$ denote the marginal period of charging and u_t^{crit} the critical level of charging for which the marginal period changes. Then, the optimal charging policy is,*

$$u_t = \begin{cases} \min\{1, \frac{x^{max} - x_t}{L}, u_t^{crit}\}, & for \quad p_t < \mathbb{E}\, p_{t+i}(1-\rho^{loss})^i; \\ 0, & for \quad p_t = \mathbb{E}\, p_{t+i}(1-\rho^{loss})^i; \\ \max\{-1, -\frac{x_t - x^{min}}{L}\}, & for \quad p_t > \mathbb{E}\, p_{t+i}(1-\rho^{loss})^i. \end{cases} \tag{6}$$

The proof of Proposition 3.1, as well as a detailed illustration of the marginal period i and u_t^{crit}, is given in Appendix 9. In case the storage constraint $(5, sto)$ is not binding, the marginal period i denotes the most profitable period during which future discharging associated with the charging decision would happen. The comparison informs the load operator about whether a charging decision should be taken. In case that constraint $(5, sto)$ is binding, the same charging or discharging decision in t can be carried out in alternative periods. In this case, i denotes the future period to which the dispatch decision in t could be most profitably postponed.

If the price in t is smaller than the expected price in the marginal period $t+i$, discounted by storage losses ρ^{loss}, the storage charges. If it is equal to the expected price in the marginal period $t+i$, charging is indeterminate and we determine that no action takes place. If the price in t exceeds the future price in $t+i$, the storage discharges. The amount of charging/discharging is exercised at the maximum feasible charging rate, unless it is not profitable to do so and charging is constrained to u_t^{crit}.

For the SO, we can infer the following implications from the optimal charging policy of Proposition 3.1. First, the charging response cannot be increased by higher prices because storage operation is a binary decision. Charging/discharging is discontinuous and occurs at the maximum feasible charging rate once the condition for charging/discharging is met. If the price change is not altering this binary decision, the storage response is inelastic. Second, even if p_t has already been decided upon and announced, the SO can still alter the charging schedule in t by changing the magnitude of future prices, in particular $p_{t+\Delta t}$. For instance, the SO might consider this necessary if new information on upcoming congestion reveals that a later storage response would be more valuable from a system perspective. In that case, the SO can set a higher $p_{t+\Delta t}$ such that an originally forecasted discharge in t does not occur and the load operator subsequently re-optimizes. In this case, the new price $p_{t+\Delta t}$ changes the marginal period of charging/discharging. Finally, while time-interdependencies and, as a result, state dependence generally limits the response of storage, higher x^{max} increases the response of storage operators as, *ceteris paribus*, the probability that the upper storage limit binds decreases.

3.2.2 Interruptible Load

An interruptible load follows a specified load profile $\overrightarrow{L}$ that can be interrupted and restarted at a later point in time. The discrete control variable $u_t \in \{0, 1\}$ describes the activity of the interruptible load. The state x_t equals the aggregated energy consumed until the beginning of period t; the state coupling is $\rho = 0$. $L(x_t)$ is the state-dependent upcoming demand. Together, these definitions consolidate into the transition function for interruptible loads,

$$x_{t+1} = x_t + L_t(x_t)u_t, \ \forall t \in \{0, 1, ..., T-2\}. \tag{4, inter}$$

Accordingly, the aggregate past consumption x_{t-1} corresponds to the expression $x_{t-1} = \sum_{\tau=t^{start}}^{t-1} L(x_\tau)u_\tau$. To provide an example, if the respective load vector is given by $\overrightarrow{L} = [1, 2, 1]$, aggregate consumption reaches $x_t = 3$ after two periods of dispatch and, accordingly, the next dispatch in $t + 1$ will be $L(x_t = 3) = 1$. L^j denotes the jth element of the load vector. b_t reflects additional costs associated with a dispatch of the load in t, e.g. costs for shifting the load. Additionally, t^{start} marks the earliest starting time and t^{end} the latest time of dispatch. Since interruptible loads must be executed to completion, the following load-specific constraint set $\mathbb{C}_{inter}$ applies,

$$x_{t^{start}} = 0, x_{t^{end}} = \sum_j L^j. \tag{5, inter}$$

The optimal dispatch policy is described by Proposition 3.2.

Proposition 3.2 *Let i denote the marginal period of dispatch, i.e. when the load would alternatively be dispatched, and t' be the first period for which an optimal dispatch starting in t versus i coincide again. Let, furthermore, Δt^{crit} be the residual number of dispatch periods necessary to comply with the terminal condition, i.e. $x_{t^{end}} = \sum_j L^j$. Then, the optimal dispatch policy is,*

$$u_t = \begin{cases} 1, & for \quad \mathbb{E}\sum_{l=0}^{t'-i-1}[(p_{t+l} + b_{t+l}) - (p_{i+l} + b_{i+l})]L^{j_t+l} \leq 0; \\ 1, & for \quad t = t^{end} - \Delta t^{crit} + 1; \\ 0, & for \quad \mathbb{E}\sum_{l=0}^{t'-i-1}[(p_{t+l} + b_{t+l}) - (p_{i+l} + b_{i+l})]L^{j_t+l} > 0. \end{cases} \tag{7}$$

The proof can be found in Appendix 10. The policy distinguishes three cases. First, component L_t^j of the interruptible load will be dispatched in t when the expected overall cost is smaller than the cost when postponing it to the marginal period i, i.e. the expected value of the cost difference is negative. Cost differences can result from a price effect or a combined price and quantity effect. A price effect occurs when an element of the load vector L^j is operated in a different period $t + l$ in which another price p_{t+l} and additional dispatch cost b_{t+l} apply. A combined price and quantity effect occurs when, due to a load shift, an element L^j of the load vector of different magnitude is active in $t + l$. In this case, the cost changes are weighted with the different elements of the load vector. For instance, the load might be partially deployed in an expensive period if it allows another larger part of the load to move into lower-cost periods. Therefore, prices, the load vector, and additional dispatch costs are all relevant for the final dispatch decisions. Second, if the remaining time in the potential dispatch window is just sufficient to realize the full load vector, the interruptible load must be dispatched without any further cost considerations. Third, if the expected cost of dispatching in t is higher than in i, the dispatch will get delayed.

For the SO, we can infer the following implications from the optimal dispatch policy. First, unlike for storage, the dispatch or load shifting decision of an operator is not only dependent on prices but also on the individual load vector $\vec{L}$ and the additional dispatch costs b, which are unknown to the SO. Second, load shifting for a given price vector is a binary decision. The achieved load response depends on the load vector $\vec{L}$, is discontinuous, and is restricted by time-interdependency. However, even though load shifting is a binary decision, different prices can lead to different responses if the load vector is not constant. Third, despite additional dispatch costs $b_{t+\Delta t}$ being unknown to the SO, by increasing DR prices $p_{t+\Delta t}$, the SO will more likely receive a response from an individual load operator. Furthermore, when additional dispatch costs are heterogeneous across loads, we can expect more interruptible loads to respond when the price is higher. Therefore, the SO can generally increase the load response of the aggregate in $t+\Delta t$ by increasing the price $p_{t+\Delta t}$. Finally, if price p_t has already been announced but new information on congestion arrives, the SO can still manipulate dispatch in t by changing the price in $t+\Delta t$, incentivizing the load operator to anticipate load to or postpone load after t. However, with time progressing, it is more likely that state dependence binds and the load dispatch cannot be anticipated or that the constraint $t = t^{end} - \Delta t^{crit} + 1$ is binding and the load cannot be postponed anymore.

3.2.3 Non-Interruptible Load

In contrast to interruptible loads, non-interruptible loads cannot be stopped once they have been started. The definition of variables and parameters of the non-interrup-tible load is identical to the previously described interruptible load. The same transition function applies,

$$x_{t+1} = x_t + L_t(x_t)u_t \ , \forall t \in \{0, 1, ..., T-2\}. \qquad (4, ninter)$$

In the set of load-specific constraints, an additional non-interruptibility constraint applies. The set of contraints $\mathbb{C}_{ninter}$ is, therefore,

$$x_{t^{start}} = 0, x_{t^{end}} = \sum_j L^j, \tag{5, ninter a}$$

$$u_t \leq u_{t+1} + \frac{x_t}{\sum_j L^j} \ , \forall t \in \{0, 1, ..., T-2\}. \tag{5, ninter b}$$

Because of the non-interruptibility constraint Equation $(5, ninter\, b)$, the only decision is the first dispatch. The optimal dispatch policy is described by Proposition 3.3.

Proposition 3.3 *The optimal dispatch policy for non-interruptible loads is,*

$$u_t = \begin{cases} 1, & for \quad \mathbb{E}\sum_{l=0}^{|\vec{L}|-1}[(p_{t+l} + b_{t+l}) - (p_{i+l} + b_{i+l})]L^l \leq 0; \\ 1, & for \quad t = t^{end} - \Delta t^{crit} + 1; \\ 0, & for \quad \mathbb{E}\sum_{l=0}^{|\vec{L}|-1}[(p_{t+l} + b_{t+l}) - (p_{i+l} + b_{i+l})]L^l > 0. \end{cases} \tag{8}$$

This is a special case of Proposition 2. As the load cannot be interrupted, cost changes need to be considered over the whole load vector, i.e. $j_t = 0$ and $t' - i = |\vec{L}|$. If cost changes are negative or zero, the load will be dispatched starting with t. If not, dispatch will be postponed. Also, the load must be dispatched if $t = t^{end} - \Delta t^{crit} + 1$.

Proposition 3.3 enables implications similar to interruptible loads regarding the optimal dispatch policy. Additionally, we see that the dispatch decision always depends on the expected cost until the final dispatch period as the load cannot be interrupted anymore once started. Therefore, if notification intervals are short, load operators may not be able to consider future price changes anymore because load operator may have already decided to start executing their loads based on their expectations about future prices. These dispatch decisions might turn out to be sub-optimal from a system perspective. As a consequence, longer notification intervals may be better to enable non-interruptible loads to respond and support the system.

3.2.4 Elastic Load

In addition to the flexible loads described by the unified framework, we consider elastic loads. Elastic loads solely respond 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 (2020b). The dispatch u_t represents the scaling factor of load $L_t(p_t^b)$ at the base price. We assume that the load can be shut down completely or increased twice compared to the baseline, and that no additional costs apply for exercising this flexibility. The dispatch is then given by Definition 3.4.

Definition 3.4 *Let ϵ describe the price sensitivity of a load, i.e. the relative price increase (decrease) for which the load is reduced to zero (doubled). Then, the dispatch policy of an elastic load is,*

$$u_t = \begin{cases} 0, & \text{for } \frac{p_t - p_t^b}{p_t^b} \geq \epsilon; \\ 1 - \frac{(p_t - p_t^b)/p_t^b}{\epsilon}, & \text{for } \epsilon \geq \frac{p_t - p_t^b}{p_t^b} \geq -\epsilon; \\ 2, & \text{for } -\epsilon \geq \frac{p_t - p_t^b}{p_t^b}. \end{cases} \tag{9}$$

If the relative price increase is at or above price sensitivity ϵ, the dispatch factor u_t decreases to a level of zero. If it decreases to or below $-\epsilon$, the load doubles. In between, u_t is linear in the price p_t and demand within this price interval can be determined by linear interpolation. In particular, u_t equals 1 for $p_t = p_t^b$, i.e. given the base price.

We can infer the following implications from Definition 3.4. First, the SO can generally achieve a larger response in $t + \Delta t$ by choosing higher prices $p_{t+\Delta t}$. Second, the dispatch of the elastic load is independent of prices in other time periods. The SO can therefore optimize prices for each time period separately. Third, the dispatch of elastic loads is independent of the notification interval Δt.

4 Solution Using Deep Reinforcement Learning

As the policy functions make clear, 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) prices in periods $\tau \geq t$. Equivalently, the aggregated system follows a Markov decision process as it can be derived as a summation of individual load operators' behavior (as previously stated in Equation (2)). The system state in t can then be characterized by the state vector of all participating loads $\vec{x}_t$.

While often the SO may have some assumptions about the existent load composition; in the worst case, no such information will be available at all. Reasons for this might be a lack of information system infrastructure or privacy concerns. We assume such a worst case scenario in which the SO is only able to observe the realized aggregate load. As a suitable solution approach, we therefore propose a model-free Deep RL approach to approximate the optimal DR policy.

4.1 Definition of States, Actions, and Rewards

Deep RL explores an environment by observing the states visited and the rewards achieved by the actions chosen. At each time step, the RL agent observes the state of the system, picks an action, and considers the reward to update beliefs. The collected experience is used to continuously update the estimated value of an action and improve the policy.

For our setup, we choose the following specifications. We characterize the *state* s_t by the hour of the day, the four net prices prior to $t + \Delta t$, and the presence of congestion in $t + \Delta t$. We consider this to be reasonable since available loads in electricity systems typically follow a daily seasonality and, as proven by the optimal policies in Sections 3.2.1 to 3.2.4, their dispatch in t depends not only on p_t but also on the prices before and after t. Furthermore, the prices p_t picked by the SO for the upcoming period $t + \Delta t$ correspond to the *actions* of the RL agent. As *rewards* r_t, we use the negative stage-wise system operation costs from Equation (1) because our Deep RL algorithm follows a value function maximization, i.e.,

$$r_t = -cong_t \cdot |L_t(\overrightarrow{p}) - L_t^{target}| \cdot \begin{cases} P_\tau & \text{for} \quad L_t(\overrightarrow{p}) - L_t^{target} \geq 0; \\ f_t & \text{for} \quad L_t(\overrightarrow{p}) - L_t^{target} < 0. \end{cases} \tag{10}$$

4.2 Deep Deterministic Policy Gradient

We use the 1DDPG algorithm as proposed by Lillicrap et al. (2015). This algorithm optimizes a policy by updating it based on the gradient of the value function. We use a learning rate of 5e-4 and 5e-3 for the actor and the critic, respectively. Further details on the implementation can be found in Appendix 11.

5 Numerical Experiments

In Section 5.1, we evaluate the profitability and robustness of our RL-based DR program for a variety of parametrizations. In Section 5.2, we analyze the load behavior, in particular, we quantify the impact of the state dependence and load composition on the DR potential.

5.1 Evaluation of the DR program

We first describe the setup of our experiments. In Section 5.1.2, we evaluate the performance of our variable DR price policy under varying system parametrizations. In Section 5.1.3, we analyze the learning behavior of the Deep RL approach. We further investigate the impact of the notification interval for the response of the loads in Section 5.1.4 and conclude by analyzing the limits of DR in handling congestion in Section 5.1.5.

5.1.1 Experimental Setup

In the following section, we describe the setup of our experiments in correspondence with Table 3, which summarizes the parameter settings.

Parameter	**Specification**
System	
$Prob(cong_{t+1} = 1 \mid cong_t = 1)$	0.3
$Prob(cong_{t+1} = 0 \mid cong_t = 0)$	0.8
Relative load reduction target ΔL	20 %
Congestion forecast quality q_t	100 %
Base price p_t^b	1
DR prices p_t	{1.1,1.2,1.3,1.4,1.5}
Penalty P_t	10
Elastic loads	
Number of loads $\lvert N_{el} \rvert$	10
Load profile $L(t)$	step function
Elasticities ϵ	$\{0.3, 0.4, ..., 0.7\}$
Load share (of peak load)	0.333
Storage	
Number of loads $\lvert N_{sto} \rvert$	5
Charging rate L	{5%, 10%}
Ratio $x^{max} : L$	{1:1,2:1,3:1,5:2}
Storage losses ρ^{loss}	{0.9, 1.0}
Storage share (of peak load)	0.1
Interruptible and non-interruptible loads	
Number of loads $\lvert N_{inter} \rvert$, $\lvert N_{ninter} \rvert$	10 (each)
Maximum length of loads $\lvert L(x_t) \rvert$	4
Flexibility window Δt^{flex}	{2,4,6}
Additional dispatch cost b_t	{0.1,0.2,0.3,0.4}
Load share (of peak load)	0.333 (each)

Table 3: Default parameters for the numerical experiments

We model the occurrence of congestion in the local electricity system as a Markov chain, with transition probabilities $Prob(cong_{t+1} = 1 \mid cong_t = 1)$ for the transition between congestion and $Prob(cong_{t+1} = 0 \mid cong_t = 0)$ between non-congestion states. We parametrize the target load L_t^{target}, so that it constantly undercuts the load level at the base price $L_t(\vec{p}^b)$ by the same amount, i.e. $L_t^{target} = L_t(\vec{p}^b) - \Delta L \cdot \max_t L_t(\vec{p}_t^b)$. We refer to ΔL as the *relative load reduction target*, in relation to the peak load $\max_t L_t(\vec{p}_t^b)$. We furthermore assume perfect forecasting capabilities of the SO, i.e. $q_t = 100\%, \forall t \in \{0, ..., T-1\}$, where the

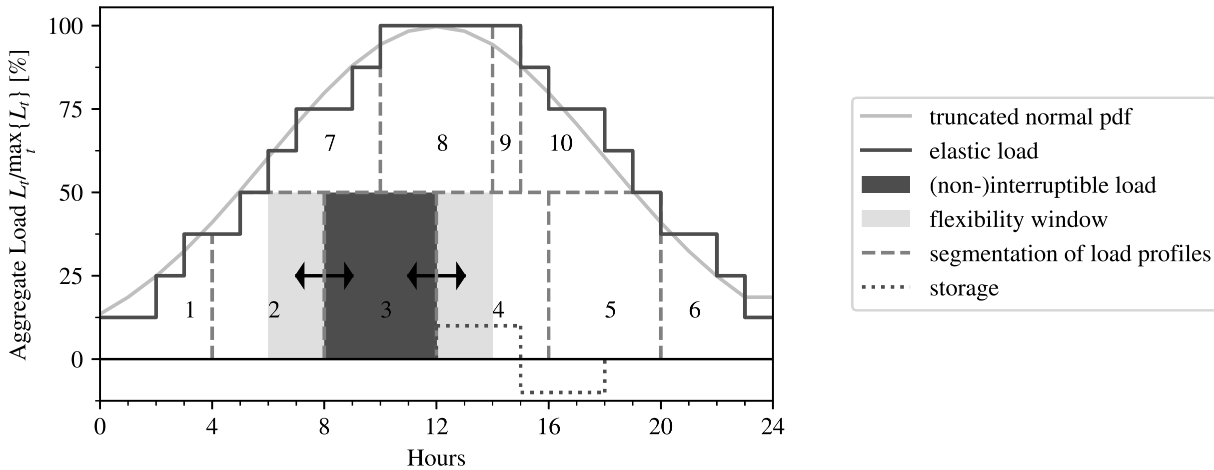


Figure 3: Generation of load profiles for interruptible loads

forecasting quality is characterized by the probability of correctly classifying a period as a congestion or non-congestion period. We normalize the time-independent base retail price to $p_t^b = 1.0, \forall t \in \{0, ..., T-1\}$, and allow for price increases in steps of ten percentage points from $p^{min} = 1.0$ to $p^{max} = 1.5$. We set the penalty for not achieving the desired load reduction to $P_t = 10, \forall t \in \{0, ..., T-1\}$.

The basic procedure for creating the load portfolio is illustrated in Figure 3 and described as follows. To start with, electricity systems typically exhibit a pronounced daily seasonality with regard to aggregate system load. We therefore derive the load shape from a truncated normal probability distribution function (light gray) which experiences its maximum at noon. It serves as the basis for the parametrization of all four load types.

For *elastic load*, we round the normal distribution to steps of 1/8 and use the resulting step function as the load profile $L(t)$ (dark gray). We assign price sensitivities ϵ between 0.3 and 0.7, in steps of 0.1, for ten elastic loads in total. For the *interruptible loads*, we slice the step function of the elastic load horizontally at 50% and randomly break the result in shorter load blocks of four hours, if possible (dashed lines, numbered blocks). This results in ten single load blocks as illustrated in Figure 3. During the core activity period (as illustrated in dark gray for one sample load block), additional dispatch costs b_t are zero. Additionally, we include a flexibility window of length Δt^{flex} of length $\{2, 4, 6\}$ which is symmetrically distributed at the start and the end of possible load activity or ends with the start or end of a day (arrows and area lightly shaded in light gray). Within this window, we parametrize additional dispatch costs $b_t > 0$ with values from the set $\{0.1, ..., 0.5\}$ which are identical throughout the flexibility window of each load. The generation of the non-interruptible loads follows the same procedure. The groups of interruptible, non-interruptible, and elastic loads are scaled such that each group represents 33.3% of aggregate load at the base price. Eventually, for *storage*, we specify five storage units according to typical volume-storage ratios and with losses $\rho^{loss} = \{0.0, 0.1\}$. We assume the aggregate charging rate to equal 10% of the aggregate peak load (dotted line), i.e. $\frac{\sum_{N^{sto}} L_n}{\max_t(L(\cdot))} = 10\%$.

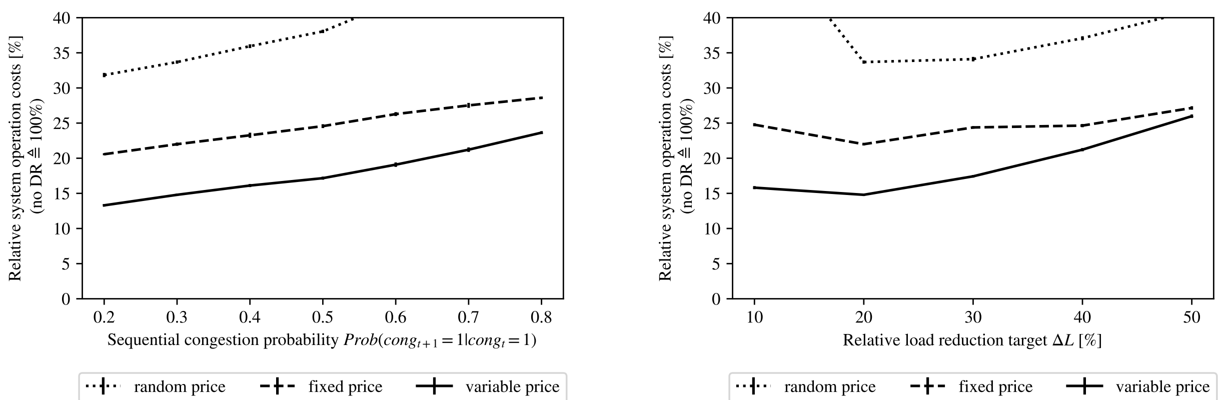
With regard to expectations about future prices beyond the notification interval, we model all load operators to be 80% myopic, i.e. they expect the latest observed price $p_{t+\Delta t}$ to continue into the future with a probability of 80% or to revert with a probability of 20% to the average DR price (if previously $\mathbb{E}\, cong_{t+\Delta t} = 0$) or base price (if previously $\mathbb{E}\, cong_{t+\Delta t} = 1$). The resulting exact load specifications are presented in Appendix 12. They remain unchanged across the following numerical experiments.

For each parameter configuration, we perform 20 separate runs for both the training and the evaluation, as the occurrence of congestion and learning are stochastic. While the size of the time steps can be arbitrarily set, we assume time steps corresponding to one hour. For each run, we train our policy over a fixed number of 5,000 time steps (approximately seven simulation months). At each time step, we re-train both the actor and the critic neural networks. We evaluate the trained actor on a separate test time series of 2,190 hours (approximately three simulation months) for each run.

5.1.2 Cost-Effectiveness of the DR Program

Default load composition. Using the previously described default parametrization as a starting point, Figure 4a shows the system operation costs for varying congestion probabilities under our variable DR price policy optimized by Deep RL ('variable price'). We compare our results to a fixed ('fixed price') and a random DR price policy ('random price'), serving as benchmarks. Under the fixed price benchmark, the same constant price is applied during all congestion events. We identify the fixed DR price by selecting among the set of possible prices $\{1.1, 1.2, 1.3, 1.4, 1.5\}$ the price which minimizes average SO costs on the training set. For the random DR price policy, at each upcoming congestion period, we randomly draw a price from the same price set, with equal probabilities.

Figure 4: System operation costs for different congestion probabilities and relative load reduction targets



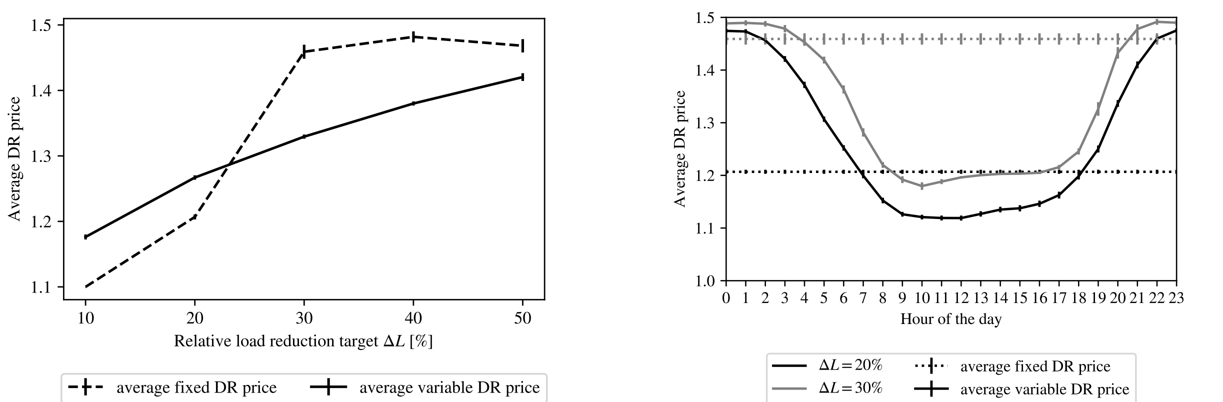
(a) Congestion probabilities

(b) Relative load reduction target

Our program, as well as both benchmarks, performs considerably better than a situation without DR (=100% of system operation costs). The random DR price policy can already achieve substantial gains by confronting loads with DR prices higher than the base price. The fixed DR price policy achieves cost of only 20.57% to 28.60% relative to the situation without DR. Our variable price DR program reduces system operation costs even further, to a level of only 13.23% to 23.64%, compared to the situation without DR. In comparison to the cost level of the fixed DR price, this corresponds to a cost reduction of 17.34% to 35.36%. In addition, our variable DR price program provides very stable results across all simulation runs. The error bars in Figure 4a represent the unbiased standard errors over all runs. In Figure 4a (as well as in all subsequent figures), they are very small and indicate that the Deep RL provides consistent results and converges to similarly attractive policies across different runs.

For the same default parametrization, Figure 4b shows the dependence of system operation costs on the relative load reduction target. Again, our program, as well as the two benchmarks, performs considerably better than a situation without DR. The variable DR price policy leads to substantially decreased system operation costs of 15.81% to 25.98%, compared to a situation without a DR program. The cost advantage compared to the fixed price policy is 4.50% to 56.65%. Furthermore, we observe that the advantage of the variable DR price decreases with an increasing relative load reduction target. This decrease is not due to an inferior performance of our program under high reduction targets. Rather, the decrease results from an exhaustion of the available DR potential: an analysis of the corresponding DR prices indicates that DR prices converge to the maximum DR price level of 1.5.

Figure 5: Average DR prices



(a) Relative load reduction target

(b) Per hour for relative load reduction targets of $\Delta L = 20\%$ and $\Delta L = 30\%$

Figure 5a shows the relation between an increasing relative load reduction target and effective DR prices. The average variable DR price rises nearly linearly from 1.2 to 1.4. In contrast, the average fixed DR price jumps from approximately 1.2 to almost 1.5 when moving from a relative load reduction target of 20% to 30%, and stays at this higher level thereafter. Figure 5b shows DR prices on an hourly basis for these two reduction targets. We can see that the variable DR prices change considerably throughout the day. During daytime, the price level is relatively low, between 1.1 and 1.2 for both relative load reduction targets, when a relatively large amount of flexible loads are active and able to respond. During the night, the same absolute response needs to be achieved with fewer active flexible loads. As a consequence, the variable DR price rises to close to 1.5 for both parametrizations. Under a fixed DR price policy, this hourly adjustment is not possible. Instead, the price must be chosen to strike a balance between day- and nighttime hours: for a relative load reduction target of 20%, sufficient DR can easily be achieved during most daytime hours. Then, a low average DR price of approximately 1.2 is sufficient to activate the necessary flexibility and minimize average expected system operation costs. However, if the relative load reduction target increases to 30%, the number of periods during which it is not easily achievable dominates and the average DR price makes a large jump to around 1.45, increasing the fixed DR price above the variable one. In summary, our analysis shows that the fixed DR price policy is optimized towards the most dominant level of scarcity in the system and is not flexible enough to dynamically respond to temporarily changing demand for DR.

Other load compositions. We now evaluate the impact of load compositions on program performance. In Table 4, each row represents the system operation costs for a different load composition and the respective results for our variable and the fixed DR price policy, again relative to a situation without a DR program. For most load compositions, the variable pricing provides a significant cost advantage and outperforms fixed DR pricing by 10% to 60%.

With regard to specific load types, high shares of *elastic loads* lead to the lowest overall system operation costs. The reason is that elastic loads can be most flexibly deployed for DR and have the least restrictive constraints among all elementary load types. Also, high shares of elastic loads are associated with a larger advantage over a fixed DR price policy than it is the case for other load types. This partially reflects the fact that learning is easier for the DDPG algorithm with a more continuous load response to prices. In contrast,

Table 4: System operation costs for varying load compositions

Load composition [%][1]				**SO costs** [%]		
Elastic Loads [%]	Interruptible Loads [%]	Non-Interruptible Loads [%]	Storage [%]	Variable DR price	Fixed DR price	Variable/ fixed
0.00	50.00	50.00	10.0	26.40	33.94	0.78
33.33	33.33	33.33	10.0	14.80	22.00	0.67
66.67	16.67	16.67	10.0	7.08	16.96	0.42
100.00	0.00	0.00	10.0	7.07	16.61	0.43
50.00	**0.00**	50.00	10.0	8.92	16.86	0.53
33.33	**33.33**	33.33	10.0	14.80	22.00	0.67
16.67	**66.67**	16.67	10.0	24.31	31.14	0.78
0.00	**100.00**	0.00	10.0	40.49	45.12	0.90
50.00	50.00	**0.00**	10.0	11.76	22.13	0.53
33.33	33.33	**33.33**	10.0	14.80	22.00	0.67
16.67	16.67	**66.67**	10.0	19.74	26.83	0.74
0.00	0.00	**100.00**	10.0	44.89	43.33	1.04
33.33	33.33	33.33	**0.0**	21.79	33.02	0.66
33.33	33.33	33.33	**10.0**	14.80	22.00	0.67
33.33	33.33	33.33	**20.0**	11.76	18.89	0.62
33.33	33.33	33.33	**30.0**	13.07	22.42	0.58
33.33	33.33	33.33	**40.0**	14.64	18.73	0.78
33.33	33.33	33.33	**50.0**	15.88	14.98	1.06

1) We 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 is held constant except for the last row block where its share is systematically increased from 0% to 50% in steps of ten percentage points.

increasing shares of *interruptible and non-interruptible loads* from 0% to 100% increases system operation costs approximately fourfold. The advantage of the variable over the fixed DR price policy remains large, reaching only 50% to 80% of the cost of the fixed DR policy. It exceeds the costs of the latter only in the extreme case of a load composition with non-interruptible loads only. Such loads can cause large load shifts and their dispatch depends on the prices of multiple subsequent periods. This makes the gradient update of actor and critic more difficult. Finally, introducing *storage* helps to decrease system operation costs; however, benefits diminish above a storage share of 30%. As the relative load reduction target for Table 4 is 20%, additional storage does not provide much additional value. The cost advantage towards a fixed DR price policy is consistent; only for a very high share of storage of 50%, the fixed price policy is advantageous. Again, we attribute this to difficulties during learning due to the large load swings storage can cause. Overall, we conclude that our variable DR program provides robust and significantly better results across different load compositions. The fixed DR price policy outperforms the variable one only in very extreme cases and only by a small amount.

5.1.3 Convergence Speed of the Deep Reinforcement Learning Algorithm

We furthermore analyze the learning behavior of the RL, particularly with regard to the ability to provide fast and stable policy recommendations. Figure 6 illustrates the learning behavior of the algorithm for 20 runs using the default scenario. The results for other parametrizations are similar and can be found in Appendix 13. The solid bold line plots the average system operation costs which can be achieved with the variable DR price policy depending on the number of training days, with one training day corresponding to 24 training periods. The thin lines show the results of the variable DR price policy for each of the 20 runs separately. The results are benchmarked by the system operation costs achieved by the fixed DR price policy, which is illustrated by the dashed line. We find that, on average, the Deep RL solution approach identifies policies performing better than an optimized fixed price policy already after a reasonable learning period corresponding to 25 days. The best exploration run provides results consistently better than the fixed price policy after a learning period corresponding to 13 days, the median policy after 23.5 days, and the worst after 45 days.

Overall, our analysis shows that the solution can be found after a reasonable learning period and that learning across runs consistently converges to an improved policy. This is important for a real-world application of the approach as the potential gains of a variable price policy must compensate for the exploration cost of finding this policy.

5.1.4 Impact of the Notification Interval

We furthermore investigate the impact of the notification interval on system operation costs. We assume that, generally, increasing notification intervals are associated with deteri-

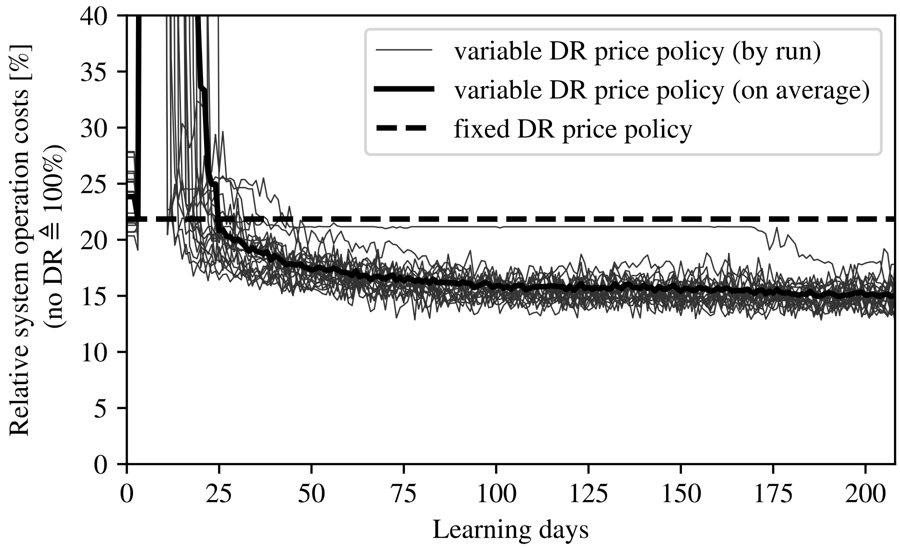
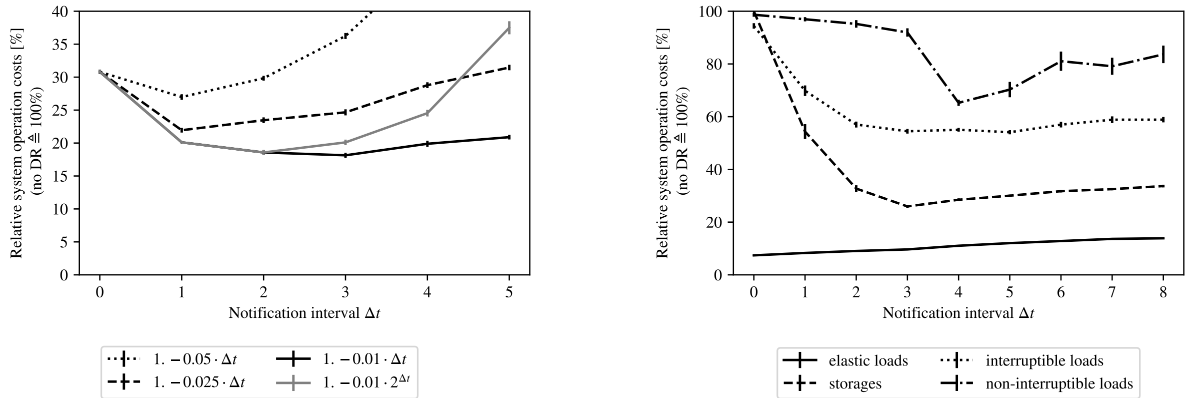


Figure 6: System operation costs depending on the training duration

orating forecasting quality. Figure 7a illustrates system operation costs for different relationships of notification interval and forecasting quality, as described in the legend. Regardless of the relationship, we see that the system operation costs improve with small increases in the notification interval in comparison to a notification in real-time. Furthermore, we observe that, for longer notification intervals, system operation costs again increase and for some functional forms even exceed the costs of a notification in real-time. For a deterioration of the forecasting quality by one percentage point per hour ahead of dispatch, the optimal notification interval is three hours while for a deterioration by five percentage points, the optimal notification interval is one hour. The most cost-effective notification interval is generally shorter the faster the forecasting quality deteriorates. In summary, this analysis demonstrates that the SO 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 an upcoming congestion (short notification intervals).

The relation between system operation costs and the notification interval is more heterogeneous when differentiated along load types. Figure 7b shows system operation costs separately for the homogeneous load compositions of the elementary load types. If the

Figure 7: System operation costs depending on the notification interval



(a) Default load composition under different relations of notification interval and forecasting quality

(b) Homogeneous load compositions of elementary load types for forecasting quality function $q(\Delta t) = 1.0 - 0.01 \cdot \Delta t$

load composition is entirely composed of elastic loads, the shortest possible notification interval leads to the lowest system operation costs because elastic loads do not face time-interdependent constraints. For the other load types, in contrast, a notification in real-time hardly produces any DR at all. For storage, the flexible load potential increases with an increasing notification interval but saturates after four hours which corresponds to the time needed to fully pre-charge and prepare for a DR event. This time span is determined by the ratio of the energy volume of the storage and the charging rate. An extension of the notification interval beyond this time span does not provide any additional value. Finally, interruptible and non-interruptible loads are the least flexible load types. Interruptible loads can respond *ad hoc* by interrupting and redispatching load, and therefore already show a response at small notification intervals. However, their response is limited by their flexibility window and, therefore, the achievable system operation costs level off for notification intervals longer than three hours. In contrast, non-interruptible loads can only respond before they are dispatched. Therefore, DR prices need to be available early on to be incorporated into the optimization of load operators. Here, the minimum system operation costs for non-interruptible loads are achieved at a notification interval of four hours.

In summary, the analyses reveal that the optimal notification interval depends on the forecasting quality and the load composition, both of which can vary considerably between real distribution systems. As a result, this design parameter needs to be tailored to the characteristics of the system of interest.

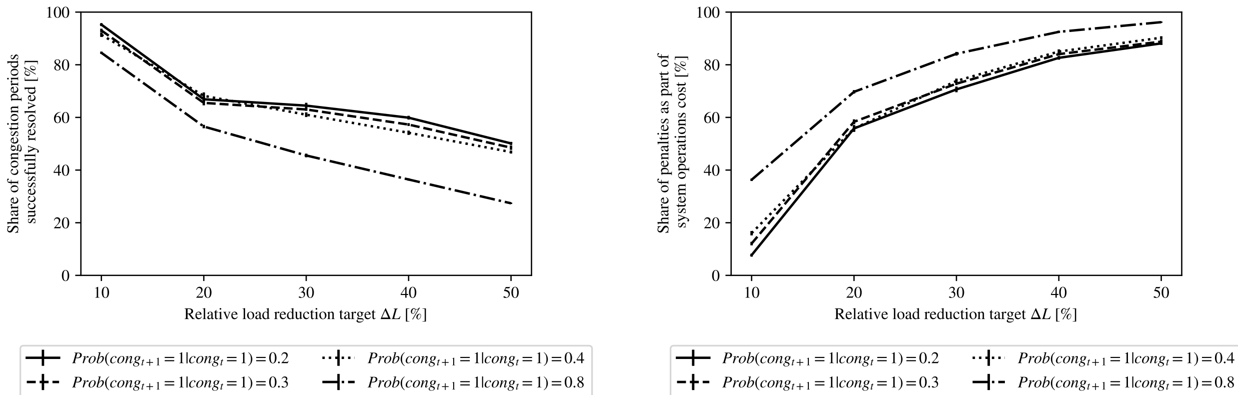
5.1.5 Limits of Demand Response

DR programs are means to provide short-term relief to a congested electricity system. However, if congestion is very frequent and large in size, DR becomes less effective and increasingly expensive compared to a long-term solution like grid reinforcements. We illustrate this below.

We begin by analyzing the impact of the magnitude of congestion, i.e. the size of the relative load reduction target. Figure 8a shows the number of periods in which DR can successfully resolve congestion, relative to the absolute number of congestion periods, for different relative load reduction targets. Each line represents a different congestion probability $Prob(cong_{t+1} = 1 | cong_t = 1)$. The figure shows that the share of successfully resolved congestion periods drops with increasing relative load reduction targets ΔL as well as increasing congestion probabilities. If the relative load reduction target is 10%, DR helps to resolve 84.55% to 95.25% of the congestion periods and proves to be a very effective method. However, if the net load reduction target increases to 50%, only 27.38% to 50.12% can be resolved. This shortcoming is also reflected in the system operation costs. Figure 8b illustrates the share of penalty payments as part of the total system operation costs. For a net load reduction target of 10%, the penalties account for only 7.68% to 36.31% of total system operation costs. For a net load reduction target of 50%, they increase to between 88.12% and 96.18%. In this case, the SO is only rarely able to resolve congestion and penalty payments account for almost all of his costs.

Second, we analyze the effect of the congestion frequency, as represented by different congestion probabilities. Both Figures 8a and 8b demonstrate that the results for congestion probabilities 0.2, 0.3, and 0.4 hardly differ; only an increase to 0.8 leads to a significant drop in performance. To understand this effect, it is important to keep in mind that 0.8 refers

Figure 8: Performance of our variable DR price program under different congestion frequencies and load reduction targets



(a) Successfully resolved congestion periods

(b) Contribution of penalty payments to total system operation costs

to the probability of experiencing two consecutive congestion periods. Higher congestion probabilities are therefore not only associated with more frequent congestion but also with more consecutive congestion periods. The drop in performance is due to the second effect: several consecutive congestion periods exhaust the flexibility potential and limit DR. When evaluating congestion magnitude and frequency together, we find that the performance is less sensitive to the congestion probability than to the relative load reduction target.

We conclude that the potential of DR programs to resolve congestion decreases with an increasing congestion probability and an increasing load reduction target. The analysis suggests that, under adverse conditions, an SO might prefer to invest in the grid infrastructure to reduce congestion frequency and size. Of course, the optimal choice between short-term (DR) or long-term measures (grid infrastructure investments) depends on additional external parameters, such as the cost of infrastructure reinforcement investments.

5.2 Evaluation of the Load Behavior

Whereas the previous section evaluated the overall system, we now focus solely on the load operators and use our unique bottom-up model of demand to analyze the DR potential. Specifically, we investigate how the DR potential is impacted by the state dependence of loads and the load composition. After presenting the setup in Section 5.2.1, we first analyze DR for a mixed load composition in Section 5.2.2 and then look at each load type separately in Section 5.2.3.

5.2.1 Experimental Setup

We use identical system and load parameters as given in Section 5.1.1. However, instead of system operation costs, we now evaluate how much DR can be achieved. For this purpose, we randomly generate a congestion time series of a length of 2,190 hours (approximately three simulation months) and evaluate DR for different combinations of notification interval and DR price. We calculate the relative DR, which is measured as the relative load reduction in comparison to the load without a DR program, i.e. $(L(p) - L(p^b))/L(p^b)$. We again average results over 20 simulation runs.

5.2.2 Behavior of a Mixed Load Composition

Figure 9 shows how much average relative DR can be achieved for different combinations of notification intervals Δt and prices p. The smallest load reduction of 7.87% is achieved with the lowest DR price of 1.1 in combination with a notification in real-time. The response increases in both price and notification interval and reaches a maximum relative DR of 54.42% for a DR price of 1.5 and the largest notification interval of four hours. The relative DR for notification intervals beyond four hours is not displayed as it hardly increases. In general, longer notification intervals increase the load response as, with earlier notice, loads have the possibility to efficiently include newly arriving information into their dispatch planning and be in the 'right' state when a response is needed. This supports our previous findings that increasing the notification interval can significantly reduce system operation costs. Similarly, a higher DR price increases the response as elastic loads respond proportionally to the price and more interruptible and non-interruptible loads with higher additional dispatch costs b adjust their dispatch.

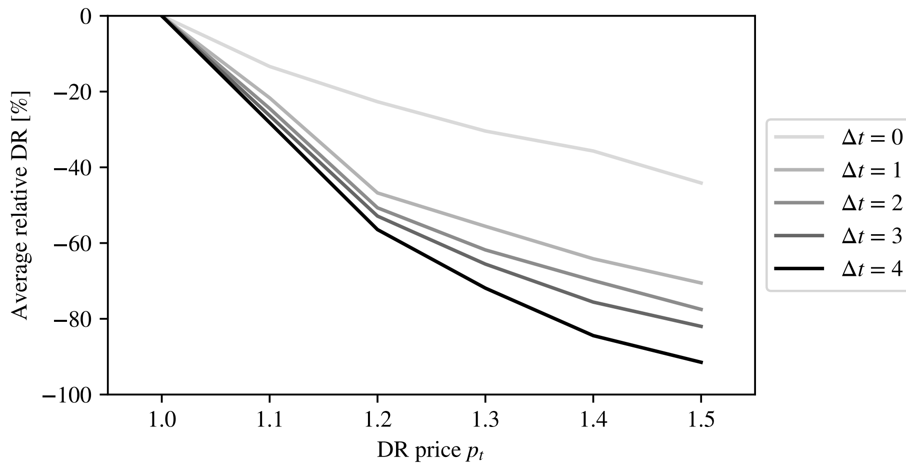


Figure 9: Average DR for different combinations of notification intervals and DR prices

Furthermore, the additional DR which can be achieved with longer notification intervals increases with prices. Approximately, the DR activity of a specific load can be associated with a minimum price and a minimum notification interval. Then, for a certain DR price, all loads with a minimum price equal or less to that DR price are responsive. This number increases with higher DR prices until, eventually, the maximum DR is achieved. The same effect occurs for increasing notification intervals. The final DR is a result of the interaction of these effects. Accordingly, if both parameters increase, the DR increases over-proportionally.

Figure 9 also illustrates that different combinations of notification interval and DR prices can lead to similar results. For instance, a DR price of 1.5 and a notification in real-time leads to a reduction of aggregate load by 28.54%. Similar load reductions can be achieved by a DR price of 1.1 and a notification interval of one hour (32.43%) or 1.2 and three hours (29.39%), respectively. This demonstrates that both design parameters are partial substitutes which can be leveraged by the SO to achieve a necessary load response. Importantly, these combinations might, however, lead to different system operation costs and might therefore not be equally valuable to the SO. Other parameters which the SO therefore has to consider are the size of the penalty payment P if DR is not sufficient to resolve congestion as well as the forecasting quality q.

5.2.3 Behavior of Elementary Load Types

We now increase the focus and evaluate how the response differs between the elementary load types. The results are displayed in Figure 10. To enable comparability, the aggregate base load in the case of no DR is identical in shape and scaled to the same level across all load types. Since the base load of storage in the case of no DR is zero, storage is scaled such that it can provide 100.00% of DR at the peak time.

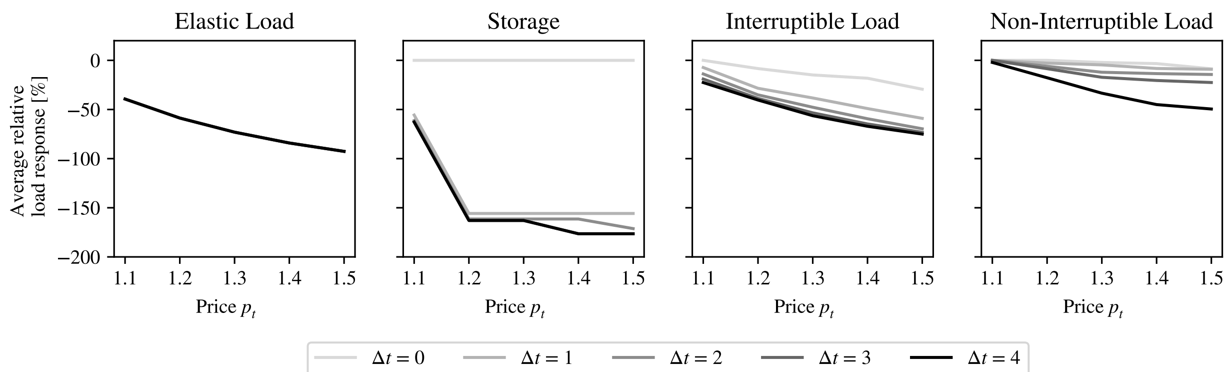


Figure 10: Average DR for different combinations of notification intervals and prices for each load type

We first analyze the differences in the response between load types. For elastic loads, a DR price of 1.1 reduces the aggregate load compared to the baseline load by 39.36%, a DR price of 1.5 by 92.62%. For storage, a price of 1.2 or higher realizes the majority of the available load reduction potential. The relative load reduction is around 150% for all notification intervals of at least one hour. Further price increases only lead to small additional load reductions, and only if combined with notification intervals of three hours or more. Then, less efficient storage units participate and increase pre-charging as the expected gains from discharging increase. For interruptible loads, relative DR increases in the price, independent of the notification interval. For instance, for a notification in real-time, the relative response increases from 0% for a price of 1.1 to 22.59% for a price of 1.5. For a notification interval of four hours, the relative response increases from 22.59% for a price of 1.1 to 75.03% for a price of 1.5. Finally, non-interruptible loads are the most constrained load type and, consequently, the least sensitive in terms of price. If the SO notifies load operators in real-time, the response is 0% for a price of 1.1 and barely increases with price. For a notification interval of four hours, the response improves and increases in price, from 0.00% to 49.51%. Overall, interruptible loads reach on average about 67.2% and non-interruptible loads 18.2% of the DR of elastic loads. These results clearly show the restrictive effect of state dependence and intertemporal constraints on DR.

We now analyze the effect of the notification interval. In contrast to previous analyses of load flexibility like Petersen et al. (2013) and Barth et al. (2018), our approach allows us to specifically analyze the impact of the notification interval on the response of different elementary load types. For elastic loads, the relative DR is equal across different notification intervals. Increasing the notification interval therefore does not increase the response. In contrast, the response considerably differs across notification intervals for the other load types. For storage, if information is only provided in real-time, no response can be achieved. Instead, providing information only one hour before dispatch already delivers most of the maximal possible response for the given DR price. Smaller additional gains can be achieved for longer notification intervals. This is especially true at higher price levels as these can compensate for storage efficiency losses during long storage durations. For interruptible and non-interruptible loads, the SO can already trigger a response when information is provided in real-time but can increase it with longer notification intervals. For interruptible loads, the marginal response (distances between curves) decreases for growing notification intervals, while the opposite is true for non-interruptible loads. The reason is that the latter can only deliver active support if the load operator did not start the dispatch yet. In our example, load vectors are four hours long. In this case, a notification interval of four hours ensures that the DR event information is always available prior to the dispatch. This way, the load can be optimally scheduled. In Appendix 14, we perform the same analysis under different load parametrizations. Overall, the analyses lead to very similar results, i.e. our results are robust to different parametrizations.

6 Conclusion

This paper proposes an online DR program which takes the specific constraints of local electricity distribution systems into account. The paper thereby fills research gaps in the literature on DR programs and on flexible load modeling. In the following section, we discuss our results and our contributions in the context of the related literature. Furthermore, we reflect on implications for management and policy and provide opportunities for future research.

6.1 Discussion of Contributions

In Section 6.1.1, we discuss our contributions to the literature on DR programs. In Section 6.1.2, we focus on the contributions which stem from our modeling of flexible loads.

6.1.1 Demand Response Programs

We propose one of the first online DR programs with variable prices. In this program, variable DR prices are determined by the SO on a rolling horizon as congestion is forecasted. Previous related programs either did not incorporate time interdependence (Yousefi et al., 2011; Mieth and Dvorkin, 2020b), have focused on a single load type (Adelman and Uçkun, 2019; Valogianni et al., 2020), or pursued other objective functions like energy cost minimization or social welfare maximization (Adelman and Uçkun, 2019; Mieth and Dvorkin, 2020b). Nevertheless, we can compare some of our results to previous findings. The DR program design by Adelman and Uçkun (2019) aims to minimize energy procurement costs for social welfare maximization without binding network constraints. They find that their dynamic pricing program can reduce the system-wide load peak by 10%. In our study, we identify much higher load reductions of up to 90%. However, in contrast to Adelman and Uçkun (2019), this decrease refers to DR participants alone since we did not include a baseload. Therefore, our results can be understood as an upper bound. For the use case of congestion management, Mieth and Dvorkin (2020b) show numerically that their program leads to (near) cost-optimal DR deployment. We cannot compare our results with the optimal solution as the decision problem of the SO is not analytically tractable. Instead, we provide a managerial perspective by comparing our results to the benchmarks of no DR and fixed DR pricing, which are similar to programs implemented in practice. Beyond Adelman and Uçkun (2019) and Mieth and Dvorkin (2020b), our results show the general advantage of a variable compared to a fixed DR price program and, furthermore, show that such a program provides significant cost advantages over a wide range of parameter settings; including congestion probabilities and relative load reduction targets. Furthermore, cost advantages are stable and reach 40% to 60% for the majority of load compositions. Only in extreme cases, based on exclusively non-interruptible loads or very high shares of storage, is our program not advantageous. To summarize, our results show that DR programs with variable pricing are a promising alternative to conventional fixed DR price programs.

Second, we provide a solution strategy to optimize DR prices for systems with non-elastic and time-interdependent demand, without requiring knowledge about the load composition. Thereby, we considerably relax assumptions of most previous work (Doostizadeh and Ghasemi, 2012; Chen et al., 2012; Valogianni et al., 2020). Similar to our approach, Mieth and Dvorkin (2020b) and Lu et al. (2018) demonstrate learning of a DR price under unknown system conditions. Mieth and Dvorkin (2020b) tailor their learning strategy to elastic loads and estimate a close to optimal DR pricing function after approximately 180 time steps. Lu et al. (2018), using Q learning, need 20,000 to 50,000 time steps to learn adequate prices for quasi-elastic loads. For a load composition of solely elastic loads, our approach requires on average 480 time steps (or 20 simulated days) to learn effective prices, which is slightly slower than Mieth and Dvorkin (2020b) but much faster than Lu et al. (2018). However, we are not restricted to this load type. Our algorithm is able to cope with different load compositions, which exhibit time-interdependencies, discontinuities, and inelasticities. Although complexity increases tremendously, our Deep RL approach needs only 600 time steps (25 days) on average to learn the optimal pricing policy. Furthermore, our strategy is functional for varying load compositions and can even be deployed with load types other than the ones described by us in the unified framework. The solution strategy is not tied to a specific load structure and only takes aggregate load as an input. This makes it applicable to a wide range of heterogeneous load compositions present in local electricity systems.

Third, we evaluate the importance of the sequential arrival of information, i.e. the impact of a notification interval for DR and the role of state dependence of loads. One of the very few academic studies that has conducted a similar analysis is Taylor and Schwarz (2000). Our results confirm the findings of Taylor and Schwarz (2000) that the optimal notification interval depends on the trade-off between greater load-side flexibility under earlier notification and the downsides of a deteriorating forecasting quality. We extend their work by analyzing the interaction between the load composition, the notification interval, and the load response. By doing so, we cannot confirm their assumption that load elasticity increases linearly with increasing notification intervals. For storage and interruptible loads, a short notification interval already enables the majority of the maximally possible response while, for non-interruptible loads, the opposite is true. We show that the specific parametrization of loads impacts the change in elasticity with time. Furthermore, Taylor and Schwarz (2000) have focused on mid-term flexibility with notification intervals of one to three days. We extend their analysis by providing a more short-term perspective with notification intervals of up to a couple of hours. By doing so, we show for the first time how the notification interval interacts with other system parameters; including DR prices, load compositions, and load parametrizations.

6.1.2 Flexible Load Modeling in DR Programs

As our first contribution regarding load modeling, we choose a bottom-up model to represent demand. By doing so, we can demonstrate that our approach performs well across different load compositions and for different load types, while previous research (Sioshansi, 2012; Wang et al., 2017; Valogianni et al., 2020) has concentrated on one load type alone. Furthermore, the bottom-up model enables us to show in detail how different load types and load parametrizations affect demand response. The most fundamental differences in the behavior of the load types are as follows. While elastic loads respond linearly to price changes, we find that storage already responds to small price changes but does not provide additional response if DR prices are further increased. In contrast, higher prices can have a positive impact on the response of interruptible and non-interruptible loads.

Second, previous work – e.g. Yousefi et al. (2011); Khezeli et al. (2017); Mieth and Dvorkin (2020a) – has mostly abstracted from state dependence. Our results show that state dependence is relevant. Moreover, we can show that the impact of state dependence increases according to how restrictive the load-specific constraint sets of load types are. Previous contributions like Petersen et al. (2013) have analyzed the impact of state dependence only in a qualitative manner. With our approach, we can quantify the impact of state dependence. For example, in the default scenario of our numerical experiment, interruptible loads can only realize two thirds of the DR provided by elastic loads, and non-interruptible loads only one fifth.

Third, our results show that our load modeling methodology based on optimal control is able to capture the impact of dynamic information arrival. Shorter notification intervals generally restrict the ability of load operators to adjust consumption and, therefore, lead to lower DR. Our bottom-up approach furthermore enables us to analyze the impact of the load composition as well as its parametrization. The response of interruptible and non-interruptible loads, for instance, can effectively be increased by longer notification intervals. For storage, in contrast, very long notification intervals do not provide additional value. These results highlight the importance of considering the dynamic information structure of the problem. Finally, we have modeled the load types in a uniform language of optimal control under uncertainty. The resulting framework extends previous frameworks (Petersen et al., 2013; Barth et al., 2018) regarding the consideration of uncertainty and the objective of load operators. The unified framework proposed by us can be a useful tool for other researchers to further study this behavior.

6.2 Managerial and Policy Implications

Our work has several implications for management and policy. First, our results demonstrate that DR programs with variable pricing are an effective approach to mitigating grid congestion. From a *managerial perspective*, we therefore recommend that SOs should work towards flexibilizing their DR programs. Unlike existing fixed DR price programs, variable pricing enables SOs to take into account changing grid scarcities throughout the day. Also, SOs should tailor their DR programs to the local characteristics of the distribution system. Locally varying load compositions, forecasting qualities, congestion probabilities, and relative load reduction targets all result in different effective parametrizations of prices and notification intervals. From a *policy perspective*, variable DR price programs can be a valuable alternative to DLMPs. Variable DR price programs have the advantage of avoiding price risk and price variability, as well as unwanted price differentiation of customers. Furthermore, they do not require extensive information system investments. Therefore, regulators should provide SOs with an appropriate framework to allow for greater flexibility of DR programs.

Second, SOs should begin to test Deep RL as a tool for congestion management. SOs typically operate in a highly cost-sensitive environment and have limited room for experimentation. Our approach is not only able to identify stable cost savings across different system characteristics; more importantly, it achieves these savings within a very short training time for a large variety of parameters and load compositions. Beyond short training intervals, the Deep RL approach has an additional advantage: it does not require information about the underlying load structure. Therefore, the approach is not subject to privacy concerns which are frequently present when dealing with granular electricity consumption data. For such algorithmic approaches to be applicable in power systems, real data only available to SOs and utilities need to be leveraged for extended testing.

Furthermore, our Deep RL-based approach has attractive characteristics from a policy perspective. In many institutional contexts, DR programs underlie regulatory approval as they temporarily increase prices for consumers. It is difficult to regulate prices and, at the same time, take into account the temporarily and locally variable requirements of the electricity system. The result of the Deep RL training process is a deterministic function (the actor network) which outputs a DR price for a given system state. Potentially, instead of specific prices, the actor network itself could be the subject of regulatory approval, providing SOs with greater flexibility while preventing abuse of their dominant position.

Finally, while fully translating the above recommendations into practice may take time, our results also yield several immediate implications. These can help SOs to refine the design of their existing time-of-use rate and critical peak pricing schedules, and other DR programs. For instance, we can derive the following general conclusions regarding the design of notification intervals: if interruptible and non-interruptible loads dominate the load composition, notification intervals should be longer and prices higher than in systems with primarily elastic loads and storage. Furthermore, if forecasting quality is high (e.g. regarding the upcoming renewable electricity generation), the SO should take advantage of longer notification intervals. Finally, if congestion becomes too frequent and necessary relative load reduction targets too high, DR is increasingly unable to resolve congestion successfully and other means, such as grid reinforcements, become more desirable.

6.3 Future Research Opportunities

While the results demonstrate the effectiveness of our approach, there are several opportunities for future research. An interesting field could be the exploration of how the different beliefs of load operators concerning future prices, as well as their forecasting abilities, impact the performance of a DR program. This could potentially provide meaningful insights on how the availability of information, forecasting tools, and automation of loads shapes DR.

Second, other cost functions for both load operators and SOs could be investigated, providing a better understanding of the interaction of DR with other pricing elements. These could, for instance, include a Demand Peak rate; which is charged based on the maximum load experienced throughout the billing period and incentivizes SOs and load operators to flatten load peaks. Eventually, beyond electricity, our approach can be applied to other markets. Of particular interest are pricing problems in environments where a central provider of a scarce good is confronted with flexible and time-interdependent demand. Relevant applications include pricing problems in job-based cloud computing (e.g. Borgs et al., 2014), supply chains (e.g. Noori-Daryan et al., 2019), ride sharing (e.g. Guda and Subramaniana, 2019), and others.

References

Aalami HA, Moghaddam MP, Yousefi GR (2010) Demand response modeling considering Interruptible/Curtailable loads and capacity market programs. *Applied Energy* 87(1):243–250.

Adelman D, Uçkun C (2019) Crosscutting Areas—Dynamic Electricity Pricing to Smart Homes. *Operations Research* 67(6):1520–1542.

Albadi MH, El-Saadany EF (2008) A summary of demand response in electricity markets. *Electric Power Systems Research* 78(11):1989–1996.

Barth L, Ludwig N, Mengelkamp E, Staudt P (2018) A comprehensive modelling framework for demand side flexibility in smart grids. *Computer Science - Research and Development* 33(1-2):13–23.

Bertsekas D (2000) *Dynamic Programming and Optimal Control*. 2nd ed. (Athena Scientific, Belmont, Mass.).

Bollinger BK, Hartmann WR (2020) Information vs. Automation and Implications for Dynamic Pricing. *Management Science* 66(1):290–314.

Bompard EF, Han B (2013) Market-Based Control in Emerging Distribution System Operation. *IEEE Transactions on Power Delivery* 28(4):2373–2382.

Borenstein S (2007) Customer Risk from Real-Time Retail Electricity Pricing: Bill Volatility and Hedgability. *Energy Journal* 28(2):111–130.

Borgs C, Candogan O, Chayes J, Lobel I, Nazerzadeh H (2014) Optimal Multiperiod Pricing with Service Guarantees. *Management Science* 60(7):1792–1811.

Boßmann T, Eser EJ (2016) Model-based assessment of demand-response measures - A comprehensive literature review. *Renewable and Sustainable Energy Reviews* 57(2016):1637–1656.

Buber T, Gruber A, Klobasa M, von Roon S (2013) Lastmanagement für Systemdienstleistungen und zur Reduktion der Spitzenlast. *Vierteljahrshefte zur Wirtschaftsforschung* 82(3):89–106.

Celebi E, Fuller JD (2007) A Model for Efficient Consumer Pricing Schemes in Electricity Markets. *IEEE Transactions on Power Systems* 22(1):60–67.

Chen L, Li N, Jiang L, Low SH (2012) Optimal Demand Response: Problem Formulation and Deterministic Case, Control and Optimization Methods for Electric Smart Grids. Chakrabortty A, Ilić MD, eds., *Control and Optimization Methods for Electric Smart Grids (Springer-Verlag, New York)*, 63–85.

Chen W, Wang X, Petersen J, Tyagi R, Black J (2013) Optimal Scheduling of Demand Response Events for Electric Utilities. *IEEE Transactions on Smart Grid* 4(4):2309–2319.

Cohen M, Lobel I, Paes Leme R (2020) Feature-Based Dynamic Pricing. *Management Science* (Articles in Advance), URL `http://dx.doi.org/https://doi.org/10.1287/mnsc.2019.3485`.

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.

Deng R, Yang Z, Chow MY, Chen J (2015) A Survey on Demand Response in Smart Grids: Mathematical Models and Approaches. *IEEE Transactions on Industrial Informatics* 11(3):570–582.

Dong C, Ng CT, Cheng TCE (2017) Electricity Time–of–Use Tariff with Stochastic Demand. *Production and Operations Management* 26(1):64–79.

Doostizadeh M, Ghasemi H (2012) A day-ahead electricity pricing model based on smart metering and demand-side management. *Energy* 46(1):221–230.

Faruqui A, Sergici S (2010) Household response to dynamic pricing of electricity: A survey of 15 experiments. *Journal of Regulatory Economics* 38(2):193–225.

Fridgen G, Saumweber A, Seyfried J, Wederhake L (2018) Decision flexibility vs. Information accuracy in energy-intensive businesses. *Proceedings of the 26th European Conference on Information Systems (ECIS 2018)*, 64.

Guda H, Subramaniana U (2019) Your Uber is Arriving: Managing On-Demand Workers Through Surge Pricing, Forecast Communication, and Worker Incentives. *Management Science* 65(5):1995–2014.

Khezeli K, Lin W, Bitar E (2017) Learning to Buy (and Sell) Demand Response. *IFAC-PapersOnLine* 50(1):6761–6767.

Kirschen DS, Strbac G, Cumperayot P, de Paiva Mendes D (2000) Factoring the Elasticity of Demand in Electricity Prices. *IEEE Transactions on Power Systems* 15(2):612–617.

Kök AG, Shang K, Yücelc S (2018) Impact of Electricity Pricing Policies on Renewable Energy Investments and Carbon Emissions. *Management Science* 64(1):131–148.

Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, Silver D, Wierstra D (2015) Continuous control with deep reinforcement learning URL `http://arxiv.org/abs/1509.02971`.

Liu Z, Wu Q, Oren SS, Huang S, Li R, Cheng L (2018) Distribution Locational Marginal Pricing for Optimal Electric Vehicle Charging Through Chance Constrained Mixed-Integer Programming. *IEEE Transactions on Smart Grid* 9(2):644–654.

Lu R, Hong SH, Zhang X (2018) A Dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach. *Applied Energy* 220(February):220–230.

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

Mieth R, Dvorkin Y (2020a) Distribution Electricity Pricing under Uncertainty. *IEEE Transactions on Power Systems* 35(3):2325–2338.

Mieth R, Dvorkin Y (2020b) Online Learning for Network Constrained Demand Response Pricing in Distribution Systems. *IEEE Transactions on Smart Grid* 11(3):2563–2575.

Mitra S, Grossmann IE, Pinto JM, Arora N (2012) Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes. *Computers and Chemical Engineering* 38:171–184.

Moazeni S, Defourny B, Wilczak MJ (2019) Sequential Learning in Designing Marketing Campaigns for Market Entry. *Management Science* (Articles in Advance), URL `https://doi.org/10.1287/mnsc.2019.3384`.

Nikzad M, Mozafari B, Bashirvand M, Solaymani S, Ranjbar AM (2012) Designing time-of-use program based on stochastic security constrained unit commitment considering reliability index. *Energy* 41(1):541–548.

Noori-Daryan M, Taleizadeh AA, Jolai F (2019) Analyzing pricing, promised delivery lead time, supplier-selection, and ordering decisions of a multi-national supply chain under uncertain environment. *International Journal of Production Economics* 209(2019):236–248.

Papanastasiou Y (2020) Newsvendor Decisions with Two-Sided Learning. *Management Science* (Articles in Advance), URL `https://doi.org/10.1287/mnsc.2019.3470`.

Parker GG, Tan B, Kazan O (2019) Electric Power Industry: Operational and Public Policy Challenges and Opportunities. *Production and Operations Management* 32(4):771.

Petersen MK, Edlund K, Hansen LH, Bendtsen J, Stoustrup J (2013) A Taxonomy for Modeling Flexibility and a Computationally Efficient Algorithm for Dispatch in Smart Grids. *Proceedings of the 2013 American Control Conference (2013 ACC)*, 1150–1156.

Schott P, Sedlmeir J, Strobel N, Weber T, Fridgen G, Abele E (2019) A Generic Data Model for Describing Flexibility in Power Markets. *Energies* 12(10):1–29.

Shoreh MH, Siano P, Shafie-khah M, Loia V, Catalão JP (2016) A survey of industrial applications of Demand Response. *Electric Power Systems Research* 141:31–49.

Sioshansi R (2012) OR Forum—Modeling the Impacts of Electricity Tariffs on Plug-In Hybrid Electric Vehicle Charging, Costs, and Emissions. *Operations Research* 60(3):506–516.

Sotkiewicz PM, Vignolo JM (2006) Nodal Pricing for Distribution Networks : Efficient Pricing for Efficiency Enhancing DG. *IEEE Transactions on Power Systems* 21(2):1013–1014.

Taylor TN, 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.

Valogianni K, Ketter W, Collins J, Zhdanov D (2020) Sustainable Electric Vehicle Charging using Adaptive Pricing. *Production and Operations Management* 29(6):1550–1572.

Wang X, El-Farra NH, Palazoglu A (2017) Optimal scheduling of demand responsive industrial production with hybrid renewable energy systems. *Renewable Energy* 100(2017):53–64.

Xu Z, Callaway DS, Hu Z, Song Y (2016) Hierarchical Coordination of Heterogeneous Flexible Loads. *IEEE Transactions on Power Systems* 31(6):4206–4216.

Yousefi S, Moghaddam MP, Majd VJ (2011) Optimal real time pricing in an agent-based retail market using a comprehensive demand response model. *Energy* 36(9):5716–5727.

Zarabie AK, Das S, Nazif Faqiry M (2019) Fairness-Regularized DLMP-Based Bilevel Transactive Energy Mechanism in Distribution Systems. *IEEE Transactions on Smart Grid* 10(6):6029–6040.

Zhang X (2014) Optimal Scheduling of Critical Peak Pricing Considering Wind Commitment. *IEEE Transactions on Sustainable Energy* 5(2):637–645.

7 Welfare Analysis

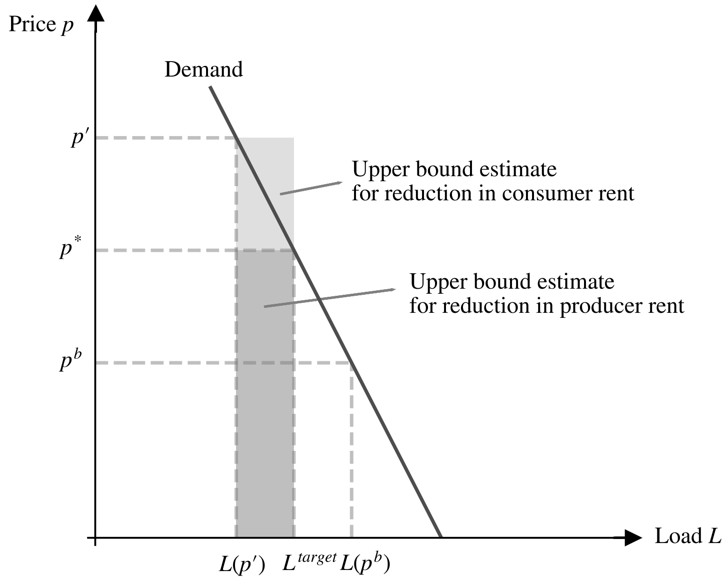


Figure 11: Welfare loss

Figure 11 illustrates the changes in welfare under a DR program. The realized load $L(p^b)$ is determined by the interaction between the demand curve and the base price p^b. However, the existence of a constraint requires the reduction of demand from $L(p^b)$ to L^{target}. To achieve this reduction, p^* would be the price of the efficient allocation, given the capacity constraint. Instead, the SO picks a price p' and potentially observes an even greater net load reduction to $L(p')$. In that case, loads with a willingness to pay between the efficient price p^* under a constraint and p' shut down which is inefficient and a welfare loss to the system. As the demand curve is not known but weakly monotonically decreasing, the consumer rent reduction can be approximated by an upper bound of $(p' - p^*) \cdot (L^{target} - L(p'))$. On the supply side, the SO does not have any information about the underlying curve. Given a lower bound estimation of supply costs being zero, producer rent reduction can be bounded by $p^* \cdot (L^{target} - L(p'))$. In sum, this results in a total estimated welfare loss of $p' \cdot (L^{target} - L(p'))$.

8 Generation-Constrained Congestion

The main text discusses DR as a measure to resolve constraints by reducing system load. However, a system can also be generation constrained and load then needs to be increased. This can be the case if, for instance, large renewable energy generation coincides with low demand, and the distribution system becomes export-constrained. In that case, the target load is higher than the expected load without a DR program, i.e. $L_t^{target} > L(p_t^b)$. To cover both generation and load constraint cases, we can generalize the objective function in Equation (1) as follows,

$$\min_{p_{t+\Delta t}} \lim_{T\to\infty} \frac{1}{T}\,\mathbb{E}_q \sum_{\tau=t}^{t+T-1} |cong_\tau|\cdot|L_\tau(\vec{p})-L_\tau^{target}|\cdot \begin{cases} P_\tau, & \text{for} \quad cong_\tau\cdot(L_\tau(\vec{p})-L_\tau^{target})\geq 0; \\ f_\tau, & \text{for} \quad cong_\tau\cdot(L_\tau(\vec{p})-L_\tau^{target})< 0. \end{cases} \tag{11}$$

Congestion $cong_t$ takes the value -1 if the system is generation-constrained. Furthermore, in that case, we define $f_t = p_t^b - p_t$. If the system is generation-constrained, effective prices under DR are lower than the base price, i.e. $p < p^b$. Therefore, the SO effectively pays load operators an amount of $|p_t - p_t^b|$ for consuming more. This aligns the SO's incentives.

9 Optimal Storage Operation

Proof 1 (Theorem 3.1) *The value function of the operator of a single storage unit reduces to,*

$$J_\tau(x_\tau) = \min_{u_\tau}\{\mathbb{E}\, p_\tau L u_\tau + \mathbb{E}\, J_{\tau+1}(x_{\tau+1})\}. \tag{12}$$

We substitute Lu_t by inserting the transfer function $Lu_t = x_{t+1} - (1-\rho^{loss})x_t$ and determine the FOC with respect to the future state x_{t+1},

$$\frac{\partial J_\tau(x_\tau)}{\partial x_{\tau+1}} = \mathbb{E}\, p_\tau + \frac{\partial\, \mathbb{E}\, J_{\tau+1}(x_{\tau+1})}{\partial x_{\tau+1}}. \tag{13}$$

In order to determine $\frac{\partial \mathbb{E} J_{\tau+1}(x_{\tau+1})}{\partial x_{\tau+1}}$ *in Equation (13), we derive a second FOC with regard to* x_τ *and update by one period,*

$$\frac{\partial J_{\tau+1}(x_{\tau+1})}{\partial x_{\tau+1}} = -\mathbb{E}\, p_{\tau+1}(1-\rho^{loss}) + \frac{\partial \mathbb{E}\, J_{\tau+2}(x_{\tau+2})}{\partial x_{\tau+2}} \frac{\partial x_{\tau+2}}{\partial x_{\tau+1}}. \tag{14}$$

Note that, due to the non-stationarity of the problem and the state constraints, we cannot leverage steady state conditions. We insert Equation (14) into Equation (13) and receive the following expression,

$$\frac{\partial J_\tau(x_\tau)}{\partial x_{\tau+1}} = \mathbb{E}\, p_\tau - \mathbb{E}\, p_{\tau+1}(1-\rho^{loss}) + \frac{\partial \mathbb{E}\, J_{\tau+2}(x_{\tau+2})}{\partial x_{\tau+2}} \frac{dx_{\tau+2}}{dx_{\tau+1}}. \tag{15}$$

If the dispatch sequence in t *hits the maximum or minimum boundary of the storage,* x^{max} *or* x^{min}*, dispatch in* t *and* $t+1$ *are substitutes and future states in* $\tau > t+1$ *will not be changed by the current dispatch decision and* $\frac{dx_{\tau+2}}{dx_{\tau+1}} = 0$*. Consequently,* $\frac{\partial \mathbb{E}\, J_{\tau+2}(x_{\tau+2})}{\partial x_{\tau+2}} \frac{dx_{\tau+2}}{dx_{\tau+1}} = 0$,

$$\frac{\partial J_\tau(x_\tau)}{\partial x_{\tau+1}} = \mathbb{E}\, p_\tau - \mathbb{E}\, p_{\tau+1}(1-\rho^{loss}). \tag{16}$$

In that case, if $\mathbb{E}\, p_\tau > \mathbb{E}\, p_{\tau+1}(1-\rho^{loss})$, $\frac{\partial J_\tau(x_\tau)}{\partial x_{\tau+1}} > 0$ *and costs are minimized if* x_{t+1} *is minimized, given the technical constraints. If that is not the case, taking the first derivative of the transfer functions gives* $\frac{dx_{\tau+2}}{dx_{\tau+1}} = (1-\rho^{loss})$ *and we can subsequently update and reinsert Equation 14 until such a boundary is hit and the sequence cut off,*

$$\frac{\partial J_\tau(x_\tau)}{\partial x_{\tau+1}} = \mathbb{E}\, p_\tau - \mathbb{E}\, p_{\tau+i}(1-\rho^{loss})^i. \tag{17}$$

i *is the marginal period of charging/discharging. The sign of* $\frac{\partial J_\tau(x_\tau)}{\partial x_{\tau+1}}$ *determines if* $x_{\tau+1}$ *should be minimized or maximized on its support interval.* ■

Example 1. For illustration, we demonstrate the solution strategy by recursion for a finite time horizon of four hours, for which $p_0 = p_1 < (1 - \rho^{loss})p_2 = (1 - \rho^{loss})p_3$ and a storage with $L = 1$ and $x^{max} = 1$. First, we specify the FOC for optimal dispatch Equation (13) for the terminal condition in $t = 3$,

$$\frac{\partial J_3(x_3)}{\partial x_4} = p_3 L. \tag{18}$$

Note that $\frac{\partial \mathbb{E} J_4(x_4)}{\partial x_4} = 0$ for the terminal condition. As $p_3 > 0$, it is therefore optimal to pick x_4 and therefore choose u_3 as small as possible to minimize cost. Due to the constraints on $x_3 \in [0, 1]$ and $L = 1$, we can derive that $u_3 = -x_3$ and $J_3(x_3) = -p_3 x_3$. Moving one step forward to $t = 2$, again, we can specify the FOC Equation 13. Inserting $J_3(x_3) = -p_3 x_3$ then yields,

$$\frac{\partial J_2(x_2)}{\partial x_3} = p_2 L - p_3(1 - \rho^{loss})L. \tag{19}$$

Note that $(1 - \rho^{loss})$ enters the equation through the transition function when updating x_3 between the end of period 2 and the beginning of period 3. Given $\rho^{loss} > 0$, we find that $p_2 > p_3(1 - \rho^{loss})$. Again, it is optimal to pick x_3 and therefore choose u_2 as small as possible to minimize cost. Given the specifications of the storage, we can derive $u_2 = -x_2$ and $J_2(x_2) = -p_2 x_2$. It follows that $x_3 = 0$ and $J_3(x_3 = 0) = 0$. We repeat the same procedure for $t = 1$,

$$\frac{\partial J_1(x_1)}{\partial x_2} = p_1 L - p_2(1 - \rho^{loss})L. \tag{20}$$

This time, $p_1 < p_2(1 - \rho^{loss})$, and $\frac{\partial J_1(x_1)}{\partial x_2} = -p_2 < 0$. Therefore, $J_1(x_1)$ can be minimized by maximizing x_2 and u_1, i.e. $u_1 = (1 - x_1)$. As a result of charging, $x_2 = (1 - \rho^{loss})$. Accordingly, the value function takes the value $J_1(x_1) = (1 - x_1)p_1 - (1 - \rho^{loss})p_2$. Finally, we solve for the first period $t = 0$,

$$\frac{\partial J_0(x_0)}{\partial x_1} = p_0 L - (1 - \rho^{loss}) p_1 L. \tag{21}$$

As $p_0 > (1 - \rho^{loss})p_1$, x_1 and therefore u_0 should be minimal. As the initial condition $x_0 = 0$ is given and $x^{min} = 0$; u_0 evaluates to 0. We can further insert into the previous equations and find that $u_1 = 1$, $u_2 = -(1 - \rho^{loss})$, and $u_3 = 0$. While it would be profitable to charge in periods 0 and 1 and discharge in 2 and 3, the constraint $x^{max} = 1$ limits the activity. As a result, the storage is charged in $t = 1$ instead of $t = 0$ and discharged in $t = 2$ instead of $t = 3$ because it is more profitable to do so.

To illustrate the marginal period i, we distinguish two cases. First, the storage operator needs to decide if it is profitable at all to charge in a certain period of t. In this *unconstrained case*, the storage constraints 5, *sto* are not binding. The cost of doing so in t is compared with the expected return in the *marginal period* i when it would be most profitable to discharge. In Example 1, it would be both profitable to charge in $t = 0$ and $t = 1$, and discharge in $t = 2$ as the most profitable discharging period. Second, given a constraint in storage volume, the storage operator needs to decide when to charge. In this *constrained case*, the storage operator needs to compare charging/discharging in t to the marginal period i when the activity would alternatively happen. For charging, periods 0 and 1 are marginal periods to each other; for discharging, these are periods 2 and 3.

The concept can easily be demonstrated if $\rho^{loss} = 0$ and one period of charging directly maps into one period of discharging. If this is not the case, the critical level of charging u^{crit} becomes relevant. If $\rho^{loss} > 0$, energy charged during one time step will be discharged in two subsequent periods, one of which might be profitable while the other will increase costs. In that case, the marginal period changes from i to i' and the optimality condition for charging might be fulfilled for i but not for i'. $u_t^{crit} < 1$ then corresponds to the share of the charging rate rate L for which the marginal period is just i and charging is favorable.

10 Optimal Dispatch of Interruptible Loads

Proof 2 (Theorem 3.2) *Because interruptible loads have an on/off control only and the cost function is not continuous nor differentiable, we instead analyze the difference in value* J_τ *from a dispatch* $u_\tau = 1$ (J_τ^d) *compared to no dispatch* $u_\tau = 0$ (J_τ^n),

$$\begin{aligned}\Delta J_\tau(x_\tau) &= J_\tau^d(x_\tau) - J_\tau^n(x_\tau)\\ &= [(p_\tau + b_\tau)L(x_\tau) + J_{\tau+1}(x_\tau + L(x_\tau))] - [0 + J_{\tau+1}(x_\tau + 0)]. \end{aligned} \tag{22}$$

We find the optimal dispatch by recursion. In the final period t^{end}, *choice is restricted because of the given load constraints. If* $x_{t^{end}} = \sum_j L^j$, *then* $u_{t^{end}} = 0$; *if* $x_{t^{end}} < \sum_j L^j$, *then* $u_{t^{end}} = 1$. *If, in the penultimate period,* $\Delta t^{crit} = 1$, *there is a choice to dispatch either in* $t^{end} - 1$ *or in* t^{end},

$$\Delta J_{t^{end}-1}(x_{t^{end}-1}) = (p_{t^{end}-1} + b_{t^{end}-1})L(x_{t^{end}-1}) - (p_{t^{end}} + b_{t^{end}})L(x_{t^{end}} = x_{t^{end}-1}). \tag{23}$$

If $p_{t^{end}-1} > p_{t^{end}}$, $\Delta J_{t^{end}-1}(x_{t^{end}-1}) > 0$ *and postponing dispatch to* t^{end} *is cost-minimizing. The opposite is true if* $p_{t^{end}-1} < p_{t^{end}}$. *Moving forward by one,*

$$\begin{aligned}&\Delta J_{t^{end}-2}(x_{t^{end}-2}) \qquad (24)\\ &= (p_{t^{end}-2} + b_{t^{end}-2})L(x_{t^{end}-2}) + J_{t^{end}-1}(x^{d,*}_{t^{end}-1}) - J_{t^{end}-2}(x^{n,*}_{t^{end}-1})\\ &= (p_{t^{end}-2} + b_{t^{end}-2})L(x_{t^{end}-2}) + J_{t^{end}-1}(x_{t^{end}-2} + L(x_{t^{end}-2})) - J_{t^{end}-2}(x_{t^{end}}),\end{aligned}$$

where $x_\tau^{d,*}$ *and* $x_\tau^{n,*}$ *denote the subsequently optimally chosen states of the load, given a dispatch* ($x_\tau^{d,*}$) *or no dispatch* ($x_\tau^{n,*}$) *in* t. *The Equation (25) compares the cost of dispatching in* $t^{end} - 2$ *and optimal dispatch after, with not dispatching in* $t^{end} - 2$ *and optimal dispatch thereafter. Again, if* $\Delta J_{t^{end}-2}(x_{t^{end}-2}) < 0$, *the interruptible load will be dispatched in* $t^{end} - 2$, *and delayed elsewise. Going forward in time yields the following general expression,*

$$\Delta J_t(x_t) = (p_t + b_t)L(x_t) + \sum_{\tau=t+1}^{t^{end}} (p_\tau + b_\tau)L(x_\tau^{d,*})u_\tau^{d,*} - \sum_{\tau=t+1}^{t^{end}} (p_\tau + b_\tau)L(x_\tau^{n,*})u_\tau^{n,*}. \tag{25}$$

As a consequence of the transition function, $x_{t+1}^{d,*} = x_t + L(x_t)$ *and* $x_{t+1}^{n,*} = x_t$. u^* *and* x^* *describe the optimal dispatch and state given the dispatch decision in* t. *The parts of the sum cancel out for all future periods* $\tau \geq t'$ *for which the dispatch profile is identical. Re-arranging and generalizing to an arbitrary alternative starting time* i *gives,*

$$\Delta J_t(x_t) = \sum_{\tau=t}^{i-1} (p_\tau + b_\tau)L(x_\tau^{d,*})u_\tau^{d,*} + \sum_{\tau=i}^{t+t'-i-1} (p_\tau + b_\tau)(L(x_\tau^{d,*})u_\tau^{d,*} - L(x_\tau^{n,*})u_\tau^{n,*}) - \sum_{\tau=t+t'-i}^{t'-1} (p_\tau + b_\tau)L(x_\tau^{n,*})u_\tau^{n,*} \tag{26}$$

i *is the marginal period for which the interruptible load would alternatively be started. Arranging the expression by load component* L^j *gives,*

$$\Delta J_t(x_t) = \mathbb{E} \sum_{l=0}^{t'-i-1} [(p_{t+l} + b_{t+l}) - (p_{i+l} + b_{i+l})]L^{j_{t+l}} \quad \blacksquare \tag{27}$$

11 DDPG Implementation

The DDPG algorithm uses two neural networks, an actor and a critic, to optimize continuous policies. The critic estimates the value function, i.e. the value of a state-action combination. The actor returns the estimated optimal policy given a specific state. While the critic is trained based on the rewards experienced for combinations of states and actions during the exploration phase, the actor is updated using the gradients of the critic at the estimated optimal action, driving the new optimal policy into the direction of the gradient of the critic.

We generally use the specifications as suggested by Lillicrap et al. (2015), in particular the structure of the neural networks for agent and critic.

Before training, we use a warm-up period of 64 time steps to collect an initial data set of states, actions, and rewards for training. For congestion periods during the warm-up period, we select actions according to a uniform distribution between possible DR prices. To allow for comparability, we restrict actions to the same discrete set of prices as in the fixed DR price benchmark. We do not yet train either actor or critic. After the warm-up, we follow Lillicrap et al. (2015) and use an Ornstein-Uhlenbeck process to explore actions, i.e. DR prices. The Ornstein-Uhlenbeck process temporally correlates chosen actions around the DR price recommended by the actor. We increase the exploration value σ to 0.4 because of the discretization of prices and to ensure wide-enough exploration for ever adapting price vectors which potentially cause large shifts in aggregate load behavior. The price in a non-congestion period is always equal to the base price and is not subject to exploration. Furthermore, we normalize prices between $[-1, 1]$ where interval boundaries are characterized by the base price and the maximum price deviation. The current state includes the hour of the day, the past four prices, and the presence of congestion. The hour of the day is represented by $sin(\frac{t}{24} \cdot 2\pi)$ and $cos(\frac{t}{24} \cdot 2\pi)$ to express that, for instance, both 11 pm and 1am are equally distant from 12am. After each time step, we store the resulting tuple of state, action, and reward to the buffer. To facilitate learning, in 50% of the non-congestion periods, we save a price different from the base price in the training database and label it with the minimum reward in the buffer, instead of storing the original tuple. By doing so, the neural networks representing actor and critic get trained on a data set with sufficient variation. This significantly stabilizes learning. At each time step, we train the actor and critic networks on a randomly drawn batch. We use a batch size of 64 and, in contrast to Lillicrap et al. (2015), higher learning rates of 5e-4 and 5e-3 for the actor and critic networks, respectively.

12 Flexible Load Parametrizations in the Numerical Experiment

Table 5: Elastic loads in the default load composition

ID	Price elasticity ϵ
1	0.1
2	0.1
3	0.2
4	0.3
5	0.4
6	0.5
7	0.5
8	0.6
9	0.7
10	0.7

Table 6: Storage units in the default load composition

ID	$L[\%]$	$x^{max} : L$	ρ^{loss}
1	10%	1:1	0.0
2	10%	5:2	0.1
3	10%	1:1	0.1
4	5%	3:1	0.0
5	10%	2:1	0.1

Each line describes one storage. The column $L[\%]$ represents the charging rate as a share of the peak load when no DR program is deployed. The charging rate is scaled such that the sum of charging rates over all storage devices corresponds to the overall storage share. The second column indicates the ratio of storage volume to the charging rate. The column on the right lists the storage losses.

Table 7: Interruptible loads in the default load composition

ID	$\vec{L}[\%]$	t^{start}	t^{end}	$\vec{b}$
1	12.5, 12.5, 25.0, 37.5	0	6	0.0, 0.0, 0.0, 0.0, 0.2, 0.2, 0.2
2	37.5, 50.0, 50.0, 50.0	3	8	0.4, 0.0, 0.0, 0.0, 0.0, 0.4
3	50.0, 50.0, 50.0, 50.0	6	13	0.1, 0.1, 0.0, 0.0, 0.0, 0.0, 0.1, 0.1
4	50.0, 50.0, 50.0, 50.0	11	16	0.2, 0.0, 0.0, 0.0, 0.0, 0.2
5	50.0, 50.0, 50.0, 50.0	14	21	0.1, 0.1, 0.0, 0.0, 0.0, 0.0, 0.1, 0.1
6	37.5, 37.5, 25.0, 12.5	19	23	0.4, 0.0, 0.0, 0.0, 0.0
7	12.5, 25.0, 25.0, 37.5	3	12	0.1, 0.1, 0.1, 0.0, 0.0, 0.0, 0.0, 0.1, 0.1, 0.1
8	50.0, 50.0, 50.0, 50.0	7	16	0.3, 0.3, 0.3, 0.0, 0.0, 0.0, 0.0, 0.3, 0.3, 0.3
9	50.0	11	17	0.3, 0.3, 0.3, 0.0, 0.3, 0.3, 0.3
10	37.5, 25.0, 25.0, 12.5	13	20	0.2, 0.2, 0.0, 0.0, 0.0, 0.0, 0.2, 0.2

The ID in the first column corresponds to the illustration in Figure 3. The column $\vec{L}[\%]$ describes the load vector, as a share of the aggregate peak load of interruptible loads when no DR program is deployed. To obtain the share in comparison to the overall system load, $\vec{L}[\%]$ must be multiplied by the share of all interruptible loads of total load. For instance, if a load component is 50%, it equals one half of the peak load of interruptible loads. Then, if the share of interruptible loads among total load is 33.33%, the contribution of this element to the total load is 16.67%. The two center columns represent the earliest start and latest end time of dispatch. The column on the right indicates the additional dispatch cost vector for the whole possible period of activity, including the period of core activity ($b_t = 0.0$) and the flexibility window ($b_t > 0.0$).

Table 8: Non-interruptible loads in the default load composition

ID	$\vec{L}[\%]$	t^{start}	t^{end}	$\vec{b}$
1	12.5, 12.5, 25.0, 37.5	0	4	0.0, 0.0, 0.0, 0.0, 0.2
2	37.5, 50.0, 50.0, 50.0	2	9	0.3, 0.3, 0.0, 0.0, 0.0, 0.0, 0.3, 0.3
3	50.0, 50.0, 50.0, 50.0	5	14	0.2, 0.2, 0.2, 0.0, 0.0, 0.0, 0, 0.2, 0.2, 0.2
4	50.0, 50.0, 50.0, 50.0	8	19	0.4, 0.4, 0.4, 0.4, 0.0, 0.0, 0, 0.0, 0.4, 0.4
5	50.0, 50.0, 50.0, 50.0	14	21	0.2, 0.2, 0.0, 0.0, 0.0, 0.0, 0.2, 0.2
6	37.5, 37.5, 25.0, 12.5	18	23	0.1, 0.1, 0.0, 0.0, 0.0, 0.0
7	12.5, 25.0, 25.0, 37.5	3	12	0.4, 0.4, 0.4, 0.0, 0.0, 0.0, 0, 0.4, 0.4, 0.4
8	50.0	7	13	0.3, 0.3, 0.3, 0.0, 0.3, 0.3, 0.3
9	50.0, 50.0, 50.0, 50.0	8	17	0.1, 0.1, 0.1, 0.0, 0.0, 0.0, 0, 0.1, 0.1, 0.1
10	37.5, 25.0, 25.0, 12.5	12	21	0.4, 0.4, 0.4, 0.0, 0.0, 0.0, 0, 0.4, 0.4, 0.4

The interpretation of the columns is analogous to Table 8.

13 Convergence Behavior of the Deep Reinforcement Learning Algorithm for Other Parametrizations

Table 9: Learning performance for different parametrizations

	Days until fixed price				Rel. SO costs [%] on last day			
Parametrization	Average	Best	Median	Worst	Average	Best	Worst	Fixed
Default	25	13	23.5	45	15.05	12.69	17.93	21.85
Default with $\Delta t = 8$	29	13	22	51	14.41	12.91	18.04	22.15
Default with $\Delta L = 50\%$	79	22	140.5	∞	65.07	57.12	72.49	67.22
Default with $Prob(cong_{t+1} = 1 \mid cong_t = 1) = 0.5$	23	11	20.5	49	17.41	15.79	20.40	24.55
Default with 100% interruptible loads	185	74	201.5	∞	47.74	42.66	55.12	49.00
Default with 100% non-interruptible loads	191	51	205	∞	57.29	52.52	62.40	49.00
Default with 100% storage	∞	156	∞	∞	27.71	18.99	46.38	20.74
Default with 100% elastic loads	20	9	17	110	7.79	6.71	10.41	19.60

The left column describes the parametrization for which the learning analysis has been run. The first four columns indicate the number of days - on average, for the best, the median, and the worst exploration run - needed to achieve consistently better results than the fixed DR price policy. The next three columns on the right side evaluate the performance of the 20 exploration runs at the end of the training period by providing relative system operation costs on average, for the best, and the worst performing run. The final column reports the average relative system operation costs under the fixed DR price policy. ∞ refers to the fact that the variable DR price policy could not be improved beyond the fixed DR price policy within the considered training period of 5,000 steps.

The results displayed in Table 9 indicate that learning durations are similar for the default load composition with different specifications of the notification interval and the congestion frequency except for very high relative load reduction targets. Regarding extreme compositions of the load portfolio, learning is very fast for elastic loads while it might take a long time for load solely composed of either storage, interruptible, or non-interruptible loads. We conclude that learning is robust and relatively fast, except for extreme parametrizations.

14 Effect of Alternative Parametrizations of Load

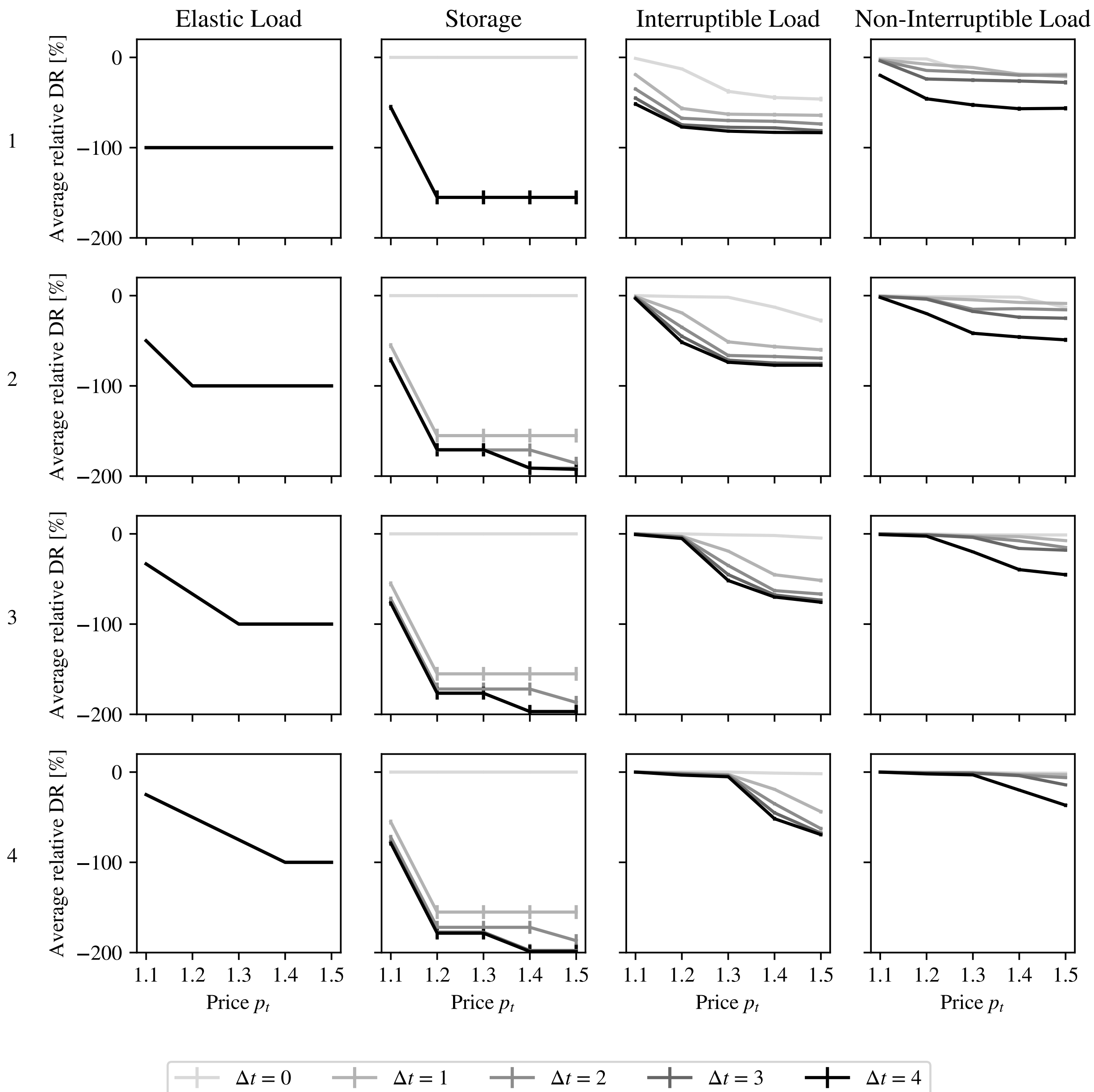


Figure 12: Average DR for different combinations of notification intervals and prices for each load type, under different parametrizations

Figure 12 shows the average relative load reductions for single load type compositions under different DR prices and notification intervals. The row numbers 1 to 4 represent the parametrization of the price sensitivity ϵ, the volume-to-charging rate ratio $x^{max} : L$, and the additional dispatch costs b_t. For example, for a parametrization of 2, the price sensitivity ϵ

is 0.2 for all loads, i.e. all elastic loads reach a 100% reduction with a price increase of 20%; all storages have a volume-to-charging rate ratio $x^{max} : L$ of 2; and all interruptible and non-interruptible loads have additional dispatch costs b_t of 0.2 outside of their core activity window.

Figure 12 confirms the central findings of Section 5.2.3. For elastic loads, the response does not depend on the notification interval but solely on the price. If the price sensitivity ϵ is low, the full flexibility potential is already achieved at low price increases. If the price sensitivity ϵ is high, the average relative load reduction increases in price and only reaches its full potential at high prices. For storage, we confirm the finding that a response can only be reached if the notification interval is longer than one hour. Furthermore, with the volume-to-charging rate ratio increasing, storage can provide slightly more response, especially when notification intervals increase. For interruptible and non-interruptible loads, Figure 12 supports our previous finding that the relative load reduction increases in both price and notification interval. In addition, we observe that, if additional dispatch costs increase, the response decreases.