Transportation Research Part C 177 (2025) 105126

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Data-driven planning of large-scale electric vehicle charging hubs using deep reinforcement learning

Karsten Schroer [a], Ramin Ahadi [a], Wolfgang Ketter [a,b,*], Thomas Y. Lee [c]

[a] University of Cologne, Cologne, Germany
[b] Erasmus University, Rotterdam, Netherlands
[c] Haas School of Business, University of California Berkeley, Berkeley, CA, USA

ARTICLE INFO

Keywords:
Digital twin
Reinforcement learning
Asset planning
Electric vehicle charging hubs

ABSTRACT

We consider the problem of planning large-scale service systems, specifically electric vehicle (EV) charging hubs (EVCHs). EVCHs are locally concentrated clusters of charging infrastructure, e.g. in large parking lots, and are often integrated with on-site generation, storage and adjacent building infrastructure. Planning such complex operational systems over a multi-year investment horizon represents a high-dimensional, dynamic and stochastic decision problem. Such planning problems typically rely on mathematical optimization frameworks which are subject to computational challenges (e.g., NP-hardness) that can limit scalability to practical system sizes. As a result, simplifying assumptions related to, for example, temporal granularity, operational detail, system size, decision horizon or stochasticity are required to achieve tractability. Modern reinforcement learning (RL) approaches, in combination with fine-grained data-driven simulation frameworks, also known as Digital Twins (DTs), may circumvent these shortcomings. We develop a scalable soft actor-critic (SAC) reinforcement learning method, that learns near-optimal EVCH configurations against a minimum cost objective. Our method uses a highly detailed DT of the EVCH environment that is bootstrapped with unique real-world sensor data from parking lots, charging stations, office buildings, and solar generation facilities, along with microscopic simulations of practical parking and charging policies. In extensive computational experiments, we provide empirical evidence that the proposed SAC RL algorithm converges closely to the global optimum (4%–15% gap) outperforming alternative popular RL approaches such as Deep Q Networks (DQN) and Deep Deterministic Policy Gradients (DDPG). We also demonstrate the superior scalability characteristic of our method to real-world problem sizes of up to 1000 charging spots. Finally, we run scenario analyses that explore the impact of user preferences and operational choices on planning decisions, thus providing actionable and novel policy guidance for EVCH planners and operators.

1. Introduction

With the proliferation of electric vehicle (EVs) arises the need for charging infrastructure that enables users to make the switch to EVs with minimal impact on lifestyle and behavior. Policymakers have traditionally assumed that users would primarily charge their EVs overnight and at home. Indeed, home charging is currently the preeminent charging use case in many markets (Lee et al., 2020; Hoover et al., 2021). As more and more consumers without access to residential charging adopt EVs, charging opportunities

* Corresponding author at: University of Cologne, Cologne, Germany.
E-mail addresses: karsten.schroer@icloud.com (K. Schroer), ahadi@wiso.uni-koeln.de (R. Ahadi), ketter@wiso.uni-koeln.de (W. Ketter), thomasyl@haas.berkeley.edu (T.Y. Lee).

https://doi.org/10.1016/j.trc.2025.105126
Received 8 July 2024; Received in revised form 23 January 2025; Accepted 13 April 2025
Available online 21 May 2025
0968-090X/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

at the workplace, at popular destinations such as supermarkets, and at fleet depots are needed (Jun and Meintz, 2018; Lee et al., 2019, 2020). We refer to the systems that afford such high-density EV charging use cases as EV Charging Hubs (EVCHs). Apart from enabling widespread EV adoption, EVCHs can also play an important systems integration role by enabling daytime charging that takes advantage of high solar energy production, which is unavailable when charging overnight (Lee et al., 2018).[1] EVCHs constitute a novel (and under-researched) operational system class with cross-system interfaces (e.g., with attached buildings or the electricity grid) and a large number of strategic and operational decision variables (size and configuration of charging stations, on-site storage, charging decisions, etc.) that result in a highly complex planning challenge (Ferguson et al., 2018).

Operations managers traditionally approach the strategic planning of operational systems like EVCHs via mathematical programming methods (e.g., He et al., 2017, for on-demand vehicle sharing service region design) or queuing models (e.g., Wang and Odoni, 2016, for last-mile delivery networks). However, the multi-stage and stochastic nature of the EVCH planning challenge makes it notoriously challenging for optimization-based methods (Powell, 2014; Hannah, 2015). Computational tractability in the traditional frameworks is only achieved at the expense of detail and scope of the planning problem. For example, shorter planning horizons, coarser temporal discretization, simplified operational detail or deterministic parameter assumptions are adopted to significantly reduce problem sizes.

This work develops a novel method that makes use of the fine-grained operational and preference data that has become abundant in this age of pervasive IoT[2] sensor technology. As such it responds to calls from the Operations Management (OM) community to incorporate such data into OM frameworks (Qi and Shen, 2018; Cohen, 2018; Choi et al., 2022) and data-driven decision support systems (Ketter et al., 2023). Specifically, we leverage fine-grained sensor data from parking lots and energy consumption and production data and combine it with high-resolution asset models and real-world operational policies into a detailed simulated environment. This environment is a close-to-exact digital representation – i.e., a Digital Twin (DT) (Choi et al., 2022; Grieves and Vickers, 2017) – of the EVCH that is to be planned. We then develop an actor-critic reinforcement learning (RL) framework that interacts with this environment to learn an optimal planning configuration policy by iterating over many simulated epochs.

Our work offers a number of contributions. Methodologically, we propose a framework for the effective use of RL in combination with large-scale data-driven simulation frameworks (i.e., DTs) for ex-ante de-risking and decision support in the design phase of service systems such as EVCHs. Our method circumvents the need for simplification and problem size reduction, among other theoretical benefits. These include (1) more realistic, data-driven modeling of stochasticity and operational detail of the EVCH, (2) computational scalability compared to mathematical optimization, and (3) flexible model setup that allows for easy evaluation of many different operational policies. In extensive simulation experiments, we show that our method achieves near-optimal EVCH planning results. We also show that it outperforms alternative candidate solution approaches such as DQN and DDPG in terms of solution speed and scalability. Finally, we make use of the flexible nature of the DT to evaluate different preference and operational regimes, thus deriving numerous novel domain-specific insights for practitioners.

The remainder of this work is structured as follows. In Section 2, we review the relevant literature to our work. We then set up and parameterize our model for data-driven planning of Electric Vehicle Charging Hubs using actor-critic RL (Section 3). In Section 4 we evaluate our model in terms of its ability to converge to the global optimum solution and its scalability and performance characteristics versus other candidate solutions, such as Deep Q-Learning. We then use the evaluated model to run comprehensive scenario analyses (Section 5) to (1) demonstrate flexibility benefits and (2) to obtain interesting and actionable policy insights. We end with a discussion of implications and contributions to theory and practice (Section 6).

2. Background

Our work draws from three main bodies of literature, which we briefly review here. First, we discuss the problem class of EVCH planning and review traditional OM planning approaches. Second, we discuss reinforcement learning (RL) methods and their potential benefits for complex, multi-stage, stochastic planning problems like EVCH planning. Third, we review the extant work on DTs and their use for OM decision support.

2.1. Electric vehicle charging hubs (EVCHs)

EV charging operations environments and use cases vary from fully distributed on-street charging, highway charging, and private home charging to charging in large-scale high-density parking lots. In this work, we focus on the latter use case, which we refer to as an EV charging hub (EVCH). EVCHs exhibit several unique features that distinguish them from other charging use cases. First, EVCHs typically represent large locally concentrated loads that may require significant local electricity grid extension making load shaping necessary (Lee et al., 2019). Second, integration with behind-the-meter loads (buildings) and generation units (PV, storage) may be desirable (Nunes et al., 2016) to reduce induced peak loads, drive sustainability and reduce costs (Ferguson et al., 2018). Third, EVCHs typically experience different user behavior compared to other charging use cases such as home charging, and this user behavior can vary substantially depending on the use case of the attached facility (workplace, mall, etc.). Fourth, siting of individual charging stations is of no concern in an EVCH context as all chargers will be located in the same space with users being largely indifferent between them. Finally, EVCHs allow for end-to-end control of the full vehicle-level parking and charging journey

[1] This is particularly relevant for energy systems with high solar energy share such as California or Germany.
[2] Internet of Things.

through what is sometimes referred to as smart EV-capable parking lots (Babic et al., 2022). This enables the assignment of vehicles to chargers and central control over the charging process. It, thus, offers new scope for optimization, e.g., by leveraging parallel or sequential use of charging equipment in an optimal manner (Ferguson et al., 2018).

We briefly review state-of-the-art OM approaches in the realms of (1) operating and (2) planning EVCHs. In terms of EVCH operations, we acknowledge the extensive work on electric vehicle charging scheduling and smart charging (see e.g., Mukherjee and Gupta, 2015 for a recent review) on which most operations-focused EVCH research is based. A notable differentiator from the traditional smart charging literature is the inclusion of building/cluster-level constraints and optimization opportunities. Early examples include (Huang and Zhou, 2015) who develop a mixed-integer optimization framework for workplace charging strategies taking into account different eligibility levels and Wu et al. (2017) who propose a two-stage energy management framework for office buildings with workplace EV charging. Nunes et al. (2016) investigate how charging processes can best be coordinated to use parking lots for EV solar-charging. Ferguson et al. (2018) propose an integrated load management approach to optimize EV charging processes for minimum cost taking into account the building base load and PV generation. A similar approach to site-level load management was implemented in practice by Jun and Meintz (2018). Finally, Lee et al. (2019) explore several optimization-driven approaches to operational issues in charging hubs. Note that the inclusion of parallel-use charging docks that allow for simultaneous charging significantly complicates that EVCH management problem. In addition to the usual charging decisions, an assignment decision of vehicles to charging stations is required. Our notion of EVCHs considers this complication.

The design/planning of EVCH systems has received less attention. EVCH design is a multi-stage stochastic decision problem that requires large decisions (e.g., the number of charging docks to be installed at each stage in the planning horizon) and small decisions (e.g., charging individual vehicles) to be taken simultaneously. Such problems are notoriously difficult and cannot be solved efficiently with standard stochastic programming or even approximate dynamic programming (Powell, 2014; Hannah, 2015). Some research resolves the ensuing complexity using simulation-based approaches. For example, in Kazemi et al. (2016) the authors use a genetic search algorithm on top of a simplified simulation model to derive the optimal size of an EV parking lot. Babic et al. (2022) also use a greedy search over a simulation of a parking lot to derive optimal infrastructure decisions. Naturally, optimality cannot be guaranteed with simple search approaches. Li et al. (2020) propose a mathematical deterministic programming framework for the joint optimization of the size and operations of a parking lot capable of 100 electric vehicles. Neither of these simulation- or optimization-based studies use high-granularity demand and/or operational data. In addition, extant EVCH design work exhibits relies on significant simplifications. For example, the studies cited here focus on a single planning period only, which reduces the problem to a single-stage planning challenge. In addition, the EVCH system scope tends to be considerably simplified (e.g., no consideration of attached building loads, single-use charging docks only, etc.).

Our work addresses several important gaps in the charging hub literature. We are the first to use detailed preference modeling in an extensive and novel set of real-world parking and charging data to ensure preference-aware sizing. In doing so, we explore the sensitivity of planning decisions to changes in user preferences, a point that has been completely neglected by existing work. In addition we consider existing building load profiles in the operations and investment decision, taking a more comprehensive view compared to previous research. Our model also allows for parallel use of charging infrastructure which can significantly boosts asset efficiency at the expense of higher operational complexity. Finally, our work has important social and sustainability implications insofar as it proposes a model for efficient provisioning of charging infrastructures that is aligned with customer preferences.

2.2. Reinforcement learning and its application to planning problems

Reinforcement learning (RL) represents a distinct class of machine learning that seeks to find an optimal policy which governs the behavior of an agent in an (emulated) environment such that a given objective is maximized (Sutton, 2019). RL relies on the Markov property, meaning that future states in a stochastic process only depend on the current state (Sutton, 2019). Popular examples of RL include an agent playing the game of Go (Silver et al., 2016) or Atari games (Mnih et al., 2013). A policy, the goal of RL, can be understood as a function that takes an observed environment state as input and returns an action given the observed state. That policy is learned iteratively by interacting with the emulator through actions and observing the effect of these actions.

RL has received significant interest as a possible approach for dynamic optimization problems (Anon, 2015). Indeed, the method boasts several potential advantages over traditional mathematical programming approaches. First, RL does not require a model but instead relies on a simulated environment to interact with and learn from. This can be a major advantage, particularly in complex multi-stage settings (like EVCH planning) where developing a mathematical model that accurately reflects the behavior of physical assets, operational policies and individual preferences is impossible or extremely hard. This also means that RL requires fewer assumptions. In mathematical programming simplifying assumptions (e.g., coarser discretization, simplified operations, removing stochasticity of inputs, etc.) are often needed to achieve tractability resulting in a problem formulation that may not reflect reality accurately enough. Second, RL is flexible and readily adapts to changes in environmental conditions. Third, RL generally deals better with the curse of dimensionality and can scale to real-sized problems well beyond the tractability limits of optimization frameworks (van Hezewijk et al., 2022). RL also has disadvantages, most notably a lack of optimality guarantee. The real-world applicability of the RL solution will also have strong dependence on the quality of the emulator that is used for training.

RL has successfully been applied to a range of operational and strategic planning problems (Gijsbrechts et al., 2022). For example, van Hezewijk et al. (2022) examine the applicability of Proximal Policy Optimization (PPO), a deep RL algorithm, to the stochastic capacitated lot sizing problem and show that the algorithm converges close to the global optimum and readily scales to problem sizes that are out of scope for traditional dynamic programming. Ahadi et al. (2023) study the charging management of shared autonomous electric vehicles using a cooperative multi-agent reinforcement algorithm to simultaneously learn optimal

scheduling and resource allocation policies. In a similar vein, Xie et al. (2023) consider a hybrid ride-hailing fleet of autonomous vehicles and conventional drivers and optimize the relocation polices using a two-sided deep RL design where the fleet operator makes central relocation decisions for autonomous vehicles and individual driver agents learn their non-cooperative relocation strategies.

There are two distinct routes for estimating the optimal policy in RL: (1) value-based approaches and (2) policy-based approaches. Value-based approaches estimate the total value associated with an action assuming the agent follows a given policy forever (e.g., the greedy policy of always selecting the action with the highest value). A value-based algorithm that has seen significant adoption is Deep Q-Learning (Gao et al., 2020). Deep Q-Learning uses Deep Q Networks (DQN) to estimate state–action values in a discrete action space. For a given state the DQN returns a Q-value for every possible action and the agent will pick a random (in the exploration phase) or the highest-valued action (in the exploitation phase). It is easy to see that DQN (and other value-based methods) can run into issues of scalability, especially if the action space is very large or even continuous, corresponding to a potentially infinite number of permutations of actions that each need to be evaluated for a given state (Dulac-Arnold et al., 2015). Policy-based methods, such as policy gradient-based algorithms can circumvent this issue and can work well in continuous action spaces (Sutton, 2019). These methods estimate the policy function directly (typically using gradient descent-based optimization) without the need to evaluate each possible state–action pair. However, they tend to be inefficient and are susceptible to local optima as well as high variance (Sutton, 2019). Actor-citric RL approaches combine value-based and policy-based approaches bringing together the benefits of both. Actor-critic algorithms consist of two main parts. First, the actor that takes decisions based on a learned policy function (policy-based). Second, the critic that determines the quality of the action using a value function (value-based). This actor-critic setup allows the actor to improve its policies more efficiently compared to pure policy-based methods. In this work, we use SAC (soft actor-critic) (Haarnoja et al., 2018), an actor-critic framework, which we adapt to work with large discrete action spaces (Dulac-Arnold et al., 2015).

2.3. Digital Twins (DTs) and their use in operations management

As mentioned, RL frameworks require a simulated environment (or simulator) to interact with and learn from. The closer this simulation comes to reality, the more likely the RL-derived solution is generalize to real-world conditions. Hence, researchers have called for the use of real-world system data to achieve more accurate representations of real-world conditions (Panzer and Bender, 2022). Digital Twins (DTs), a form of data-driven simulation, provide an attractive solution. DTs have been hailed as a disruptive trend in OM of the IoT and Industry 4.0 era (Choi et al., 2022). At its most basic level, a DT is a digital representation of a specific physical asset or system. DTs can be used as a decision support tool along the entire life cycle of that asset: from design, operations, and maintenance to disposal (Schleich et al., 2017). For the purpose of this research, we highlight several characteristics that we consider key and distinguish DTs from, e.g., traditional simulation frameworks used in OM (Boschert and Rosen, 2016) and RL. The reader is referred to Jones et al. (2020) or Cimino et al. (2019) for comprehensive reviews.

First and foremost, DTs represent a close to the real-world representation of the physical system at a granular level using high-fidelity interconnected physical models of system components (Glaessgen and Stargel, 2012). Note that this does not necessarily require identical accuracy but can also mean partial accuracy if this is sufficient for the DT to fulfill its intended use (van der Valk et al., 2020). Second, a DT is data-driven, meaning that it primarily relies on real-world operational data input acquired directly from sensors of the device and the intended application environment (van der Valk et al., 2020). It is often not the case that raw data on all required parameters is available. In such cases, synthetic data from statistical models or simulation frameworks can be used to supplement the data requirements (Sierla et al., 2018). Third, there is eventual synchronization of the DT and the physical asset. This can be achieved via one-directional (physical world to DT) or bi-directional data flows (Tao et al., 2018). DTs have been successfully applied in the use phase of operational systems (Jones et al., 2020). Examples include asset status and health monitoring (Glaessgen and Stargel, 2012), asset optimization, maintenance planning (Cimino et al., 2019), and staff training (Choi et al., 2022).

As argued previously, a core feature of DTs is their reliance on real-world sensor data for asset and process representation in the virtual world. Use of real-world environmental context data and benchmark infrastructure sensor data is also useful for design/configuration applications of assets and service systems (Attaran and Celik, 2023). Indeed, high-granularity data on many aspects of the intended application environment as well as on typical machine/process behavior is likely to already be available. This means that pre-use-phase DTs can be fed with live as well as realistic historic data (Boschert and Rosen, 2016).

In this work, we leverage the DT concept to create a high-resolution, data-driven simulation environment that resembles real-world conditions as closely as possible. Wherever available, we leverage historic sensor data to model system dynamics. For example, we use real-world charging data, parking data, building data and photovoltaic production data to achieve granular patterns of expected power production and consumption in the EVCH. For system components where real-world sensor data is not available, we rely on close-to-exact simulations of their physical behavior in line with asset specification sheets and/or research findings. For example, to model the physical properties of an on-site battery storage asset, we draw on research by Ghiassi-Farrokhfal et al. (2016) to parameterize battery charge and discharge efficiency as well as minimum and maximum (dis-)charge rates. Hence, we follow Sierla et al. (2018) to supplement our DT simulation with data from simulation frameworks to fulfill our data requirements.

We show that DTs, in combination with scalable RL algorithms, can circumvent many of the model simplifications and problem reduction measures unavoidable in traditional mathematical programming or brute-force search strategies.

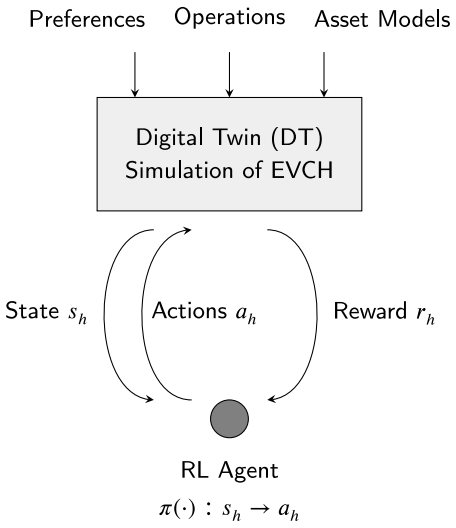

Fig. 1. Overview of Core Model Elements, Inputs and Interactions.

3. Model

We now describe our model. A conceptual overview of its core elements and their interactions is shown in Fig. 1. In this Section we describe the setup of the model starting with a definition of the planning problem including state boundaries, action space and objective followed by a description of the environment simulator (DT) and the SAC RL framework.

3.1. Defining the EVCH planning scope

We define an EVCH as an EV charging-capable parking lot, depot, or garage that is typically attached to an existing building with a given baseload. Both the building and the EVCH receive power from the same grid connection point, which is constrained to the capacity of the on-site substation. The integrated facility may have additional on-site behind-the-meter generation (e.g., photo-voltaic (PV)) and storage (e.g., Lithium-Ion battery). Crucially, charging docks can have multiple connectors that afford parallel charging of vehicles from a single charging dock. A simplified view of the EVCH components and system boundary is depicted in Fig. 2.

We formulate the EVCH configuration challenge as a feasibility problem that aims to satisfy all or a specified amount of total charging demand in the most resource-efficient manner while considering any exogenous rate, space, and total capacity constraints. The problem then becomes a cost minimization planning with the objective to jointly minimize investment costs (C^{Φ}) and operations cost (C^{Ω}) over all stages $h \in \mathcal{H}$ contained in the planning horizon while ensuring a pre-defined service level η^h (typically 100% in the following simulations and benchmarks). As we lay out the variables and parameters of our model please refer to Table 1 for an overview of nomenclature used throughout this paper.

Formally, the objective function $f(\Gamma)$ (where Γ is the system configuration) can be expressed as follows:

$$Min_{\Gamma}[C^{\Phi}(x^{i,n}_{k,h}, \delta^{Trafo}_{h}, \kappa^{PV}_{h}, \epsilon^{Bat}_{h}) + C^{\Omega}(\omega_{k,j,h}, \psi_{k,j,h,t}, \beta^{Charge}_{h,t}, \beta^{Discharge}_{h,t}, e^{G}_{h,t})] \quad (1)$$

The EVCH infrastructure decision space determining C^{Φ} extends over a large set of decision variables, which we briefly describe here. First, decisions on the charging infrastructure configuration and scale-up over the investment horizon $\mathcal{H}$ are required. We allow full flexibility regarding the type of EV charging docks (22 kW AC or 50 kW DC docks) and the number of connectors per dock (ranging from single-connector setups to up to four connectors per dock). Crucially, for charging docks with multiple connectors, we allow for simultaneous charging of EVs, meaning the rated power per dock can be shared dynamically and flexibly by all connected vehicles. This is different from the more prevalent single-server docks, which either possess just a single connector or multiple connectors that may only be operated sequentially. A multi-server setup has the advantage of higher utilization (vehicles that have completed their charging cycle do not block charging docks) (Ferguson et al., 2018). We capture EV charging infrastructure decisions via a set of binary indicator variables of form $x^{i,n}_{k,h}$, indicating whether a dock of type $i \in \{22\text{ kW}, 50\text{ kW}\}$ with number of connectors $n \in \{1, 2, 4\}$ is to be installed at candidate point $k \in K$ during planning stage $h \in \mathcal{H}$. The total number of docks and connectors is naturally bounded by the size of the facility (i.e., number of parking spaces) L. Second, the initial size and expansion pathway of possible on-site generation (PV) and/or storage assets (Li-Ion battery) must be defined. We assume that PV generation κ^{PV}_{s} can be scaled close-to continuously across all stages h in the planning horizon and that it is limited only by local facility space constraints R (e.g., roof space). In terms of on-site storage, we consider Li-Ion battery technology whose energy capacity κ^{Bat}_{s} (in kWh) can be scaled continuously over $\mathcal{H}$. Finally, a decision is required on whether, by how much and by when the existing substation capacity should be extended to accommodate the desired level of charging service. Note that substations can be purchased in standard sizes

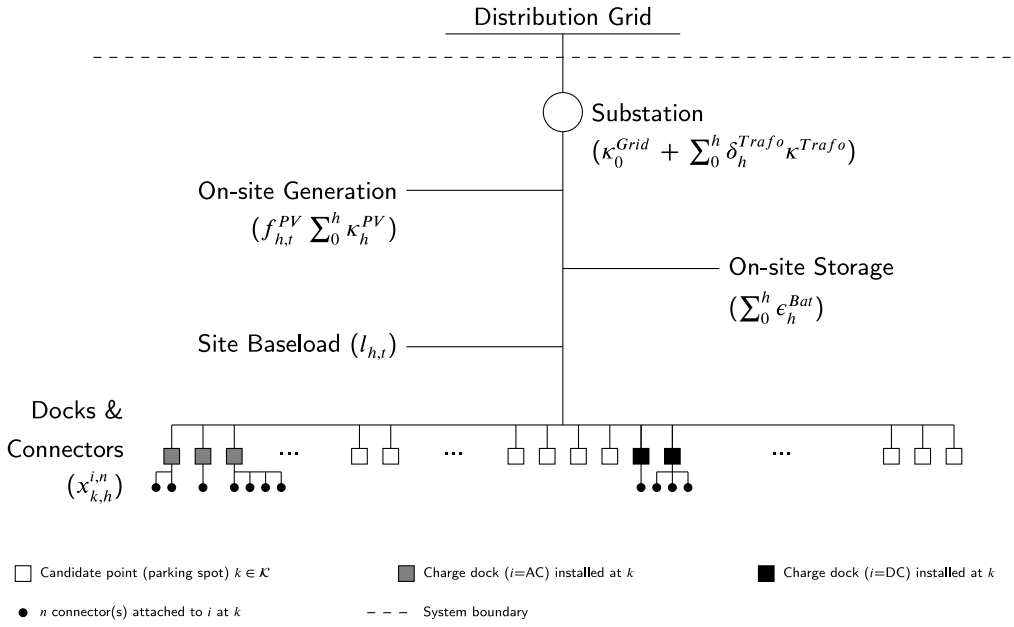

Fig. 2. EVCH Service System Layout and Asset Components in Planning Stage h.

(κ^{Trafo}) only. Consequently, the grid connection can only be scaled step-wise in multiples δ_h^{Trafo} of κ^{Trafo}, where δ_h^{Trafo} is an integer value denoting the number of transformer modules to be added to the facility's substation in state h.[3] Given the physical size of substations and the fact that local grid conditions may not allow for an unconstrained scale-up of the existing grid connection, we impose a maximum G on the final size of the substation. In sum, total investment cost over the planning horizon is determined as follows: $C^{\Phi} = \sum_{h\in\mathcal{H}}(c_h^{Trafo}\delta_h^{Trafo} + \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}}\sum_{n\in\mathcal{N}} c_h^{i,n}x_{k,h}^{i,n} + c_h^{PV}\kappa_h^{PV} + c_h^{Bat}\epsilon_h^{Bat})(1 + (|\mathcal{H}| - h)\mu^{Maint}))$. Note that this includes the maintenance costs incurred over the planning horizon, which is captured by the factor $(|\mathcal{H}| - h)\mu^{Maint}$). The parameters c_h^{Trafo}, $c_h^{i,n}$, c_h^{PV} and c_h^{Bat} are stage-dependent cost parameters that take into account expected technology cost trajectories over the planning horizon.

Underlying these higher-level infrastructure choices are smaller operational decisions, which determine the operational cost C^{Ω}. Naturally, the operational scope is constrained by the installed infrastructure highlighting the two-way interdependencies between both sets of decisions. Operations decisions focus on the assignment of a vehicle j to a connector k upon arrival (captured by $\omega_{k,j,h}$) and the periodic charging decisions over the duration of stay ($\psi_{k,j,h,t}$). Finally, the on-site battery state is controlled via $\beta_{h,t}^{Charge}$ and $\beta_{h,t}^{Discharge}$, two booleans that control the rate of charge/discharge. We consider PV generation and building baseload to be exogenous parameters that cannot be actively controlled by the EVCH operator. Given the different sources of power (battery, PV, grid), the operator also needs to decide on the power mix per each period t. This involves setting the desired energy drawn from the grid per period in each state $e_{h,t}^{Grid}$. Note that $e_{h,t}^{Grid}$ is typically accounted for based on a two-part tariff including a time-of-use-dependent energy charge $T_{h,t}^{e}$ and a monthly demand charge T_h^{p} that is a function of the maximum induced power p_h^* in that month. Operational costs are formally defined as follows: $C^{\Omega} = \sum_{h\in\mathcal{H}}(\sum_{t\in\mathcal{T}} T_{h,t}^{e} e_{h,t}^{Grid} + T_h^{p} p_h^*)$.

3.2. Parameterizing the Digital Twin simulator

We now set up the environment with which the RL agent will interact. The goal is to create a close-to-exact digital representation (i.e., a DT) of the EVCH, which is made up of three core components: (1) physical assets, (2) operational policies that define how the physical assets are operated, and (3) preference/demand characteristics, i.e., the external requests and usage patterns that the service system needs to fulfill. In this section, we lay out these three components. Note that the EVCH DT operates on a discrete-time basis. To reduce any issues/inconsistencies related to discretization, we use period lengths of just one minute. Wherever practical the DT is fed with real-world sensor data to achieve a highly accurate representation of the physical world. Simulation is used in areas where data are not available. For an overview of data sources and digitalization approaches per DT component refer to Table 5 in Appendix D.

[3] We consider any interaction effects with the upstream electrical distribution grid to be out of scope for this problem. Specifically, we assume that the distribution grid is unconstrained and able to accommodate any additional load from the EVCH, provided sufficient substation capacity (i.e., transformers) is installed for voltage regulation.

Table 1
Nomenclature.

Symbol	Description	Unit
Sets		
$\mathcal{H}$	Set of planning stages in planning horizon with index h	set
$\mathcal{I}$	Set of charging dock types with index i ($\mathcal{I} = \{AC, DC\}$)	set
$\mathcal{J}_h$	Set of unique EVs entering the EVCH during the planning period h with index j	set
$\mathcal{K}$	Set of charging dock candidate points (i.e., parking spots) with index k	set
$\mathcal{N}$	Set of charging dock connector options $\mathcal{N} = \{1, 2, 4\}$ with index n	set
$\mathcal{T}$	Set of time periods per each stage in planning horizon with index t	set
Ξ	Set of decision variables	set
Γ	Full configuration of EVCH system	set
Parameters		
$A_{j,s}$	Arrival time of vehicle j in stage h	period t
β^{max}, β^{min}	Maximum charge and maximum discharge rate of energy storage	kW
$c_h^{i,n}$	Cost per EV charging dock of type i with n connectors in stage h	USD
c_h^{Trafo}	Cost per kW of grid connection (i.e., transformer) in stage h	USD/kW
c_h^{PV}	Cost per kWp of PV in stage h	USD/kW
c_h^{Bat}	Cost per kWh of energy storage (battery) in stage h	USD/kWh
δ_j	Duration of stay of vehicle j	hours
Δ_t	Duration of a single planning period t	hours
$D_{j,h}$	Departure time of vehicle j in stage h	period t
e_j^d	Total energy requested by vehicle j over duration of stay	kWh
$\eta^{(dis)charge}$	Charge/discharge efficiency of energy storage	ratio
η^{Inv}	AC-DC inversion efficiency	ratio
η_h^{Serv}	Target service level expressed as ratio of fulfilled vs. actual demand	ratio
$f_{h,t}^{PV}$	Avg. PV load factor in period t of stage h	ratio
κ^i	Maximum power per charging dock of type i	kW
κ_0^{Grid}	Existing facility substation capacity	kW
κ^{Trafo}	Standard size of transformer that can be installed	kW
lax_j	Laxity of vehicle j	hours
$l_{h,t}$	Base load of attached facility during period t in stage h	kW
l_h^*	Maximum expected base load of attached facility in stage h	kW
M	big-M constraint (for linearization)	kW
μ^{Maint}	Cost ratio for maintenance (as share in total capital stock)	ratio
R	Maximum installable PV capacity (space constraint)	kWp
SoC^{max}	Maximum energy storage level	%
SoC^{min}	Minimum energy storage level	%
L	Space limitation in number of parking spots	count
T_h^p	Cost of induced power peak per accounting period (i.e., demand charge) in stage h	USD/kW
$T_{h,t}^e$	Cost of energy in period t of stage h as per TOU tariff	USD/kWh
$U_{j,h,t}$	Indicator of whether vehicle j is present during period t in stage h	boolean
Variables		
$\beta_{h,t}^{Charge}$	Charge rate of EVCH battery storage	kW
$\beta_{h,t}^{Discharge}$	Discharge rate of EVCH battery storage	kW
$\beta_{h,t}^{Direction}$	Indicator of whether the battery is charging or discharging	boolean
δ_h^{Trafo}	Number of additional transformers installed in stage h	integer
C^{Φ}	Total normalized investment cost for the EVCH over planning horizon	USD
C^{Ω}	Total cost of operating the EVCH over the planning horizon	USD
C^{Ψ}	Penalty for not serving charging demand	USD
$e_{j,h,t}^S$	Net energy supplied to vehicle j during period t of stage h	kWh
$e_{h,t}^{Grid}$	Net energy supplied from grid during period t of stage h	kWh
ϵ_h^{Bat}	Installed energy storage capacity in stage h	kWh
κ_h^{PV}	Installed PV capacity in stage h	kW
p_h^*	Induced max peak attributable to EVCH operations during stage h	kW
$\psi_{k,j,h,t}$	Charge rate of vehicle j connected to charging dock k during period t of stage h	kW
$SoC_{h,t}$	State variable that tracks state of charge of energy storage	kWh
$\omega_{k,j,h}$	Indicator for whether a vehicle j is connected to charging dock k in stage h	boolean
$x_{k,h}^{i,n}$	Indicator whether dock (type i, n connectors) is installed at k in stage h	boolean

3.2.1. Digital EVCH asset models

Fig. 2 provides an overview of the physical EVCH asset classes that are to be represented digitally in the DT environment. We draw on asset spec sheets along with real-world machine data to represent the physical EVCH components and the context they operate in as accurately as possible.

Local substation. We model the local substation as an integrated system consisting of transformers, circuit breakers, and other peripheral equipment that connect the site to the higher voltage levels of the distribution grid. The substation capacity is determined

by the sum of rated transformer capacities. Although typically very low, we account for transformer losses using an efficiency factor of 1-η^{Trafo}=2%.

On-site electricity generation assets (PV panels). For electricity generation, we assume photovoltaic (PV) modules, a natural supplement to EV charging hubs due to their production patterns that are highly correlated with occupancy profiles (and thus charging demand) of most parking lots. PV generation is non-dispatchable, i.e., it cannot be actively controlled. PV power is therefore consumed on-site (by EVs, battery storage, building, etc.), or fed back into the grid. Note that PV installations require DC-AC conversion via inverters. The efficiency losses of DC-AC conversion are accounted for via an inversion efficiency factor of 1-η^{Inv}=4%. We use real-world PV load factors ($f_{h,t}^{PV}$) to model PV production from the regions corresponding to the intended EVCH facility locations. Load factors are a measure of real PV panel power output as a ratio of installed capacity (κ_h^{PV}) and depend on local solar irradiation conditions.[4] PV production at time t (excluding DC-AC conversion losses) is then given by $f_{h,t}^{PV}\kappa_h^{PV}$.

Electricity storage assets. We model electricity storage as a lithium-ion battery with instantaneous ramp time. To avoid excessive battery degradation, we allow the state of charge to vary over the interval of $[5\%, 95\%]$, thus avoiding deep discharging and overcharging that are particularly strenuous for battery hardware. Setting upper and lower energy content boundaries is a common approach in storage management (Ghiassi-Farrokhfal et al., 2016). We also assume symmetric charge and discharge efficiency of $\eta^{charge} = \eta^{discharge} = 95\%$. Battery operations are simulated using what is sometimes referred to as a C/C/C model.[5] In addition, we implement typical battery constraints related to a maximum charge and discharge rate κ^{Bat} (symmetric). κ^{Bat} is dependent on the size of the battery and is set such that the battery can be charged/discharged to/from full charge within one hour.

Peripheral building. We use real-world building consumption data to model site baseload ($l_{s,t}$) that is served by the same grid connection, thus influencing total available grid capacity at any given period t (see Fig. 2). Contrary to EV loads, we assume $l_{s,t}$ to be exogenous, i.e., it cannot be dynamically managed or even curtailed. Given the absence of smart energy management hardware in most existing building stock, this is a reasonable assumption. Note that granular consumption data is widely available for commercial buildings above a certain consumption threshold since these consumer classes are typically exposed to time-of-use tariffs as well as demand charges for induced peak load. Our 1-year dataset records peak building loads and consumption at a 15 min resolution.

EV charging docks and connectors. We model two different types of charging docks that mainly differ in terms of maximum charging rates κ. Specifically, we allow for AC fast chargers with maximum charging capacity $\kappa^{i=AC}$=22 kW and DC super-fast chargers with maximum charging capacity $\kappa^{i=DC}$=50 kW. For each charger type $i \in [AC, DC]$ we allow for different connector configurations with $n \in [1, 2, 4]$ connectors per dock. We assume that κ can be shared dynamically and flexibly between all connectors per dock. This means that connected EVs can be served both sequentially and simultaneously via the same dock. Losses related to AC-DC conversion are modeled using an efficiency factor of 1-η^{Inv}=4%.[6]

3.2.2. EVCH operational policies

We also implement a range of realistic operational policies that simulate real-world operations in the DT environment. These policies are inspired by standard operational practices currently used in EV charging operations as well as recent algorithms proposed in the EV charging literature (e.g., Lee et al., 2019; Ferguson et al., 2018). Given that we allow for multi-connector charging docks with simultaneous charging capability, the initial assignment of vehicles to charging stations becomes important due to heterogeneous energy demand and flexibility characteristics. Clearly, since EVs cannot be readily relocated while parked, the initial assignment to a connector influences future available charging capacities for the EV in scope as well as for current and future arrivals that are to be served by the same charge dock.

Vehicle routing algorithms. Vehicles $j \in \mathcal{J}$ are routed/assigned to a connector k upon entry into the EVCH (captured by $\omega_{k,j,h}$). We implement two heuristic routing algorithms of varying levels of sophistication and varying information requirements.

- **Lowest-utilization-first (LUF)**: This strategy operates on a sorting basis. At each new arrival, the algorithm sorts all available docks based on free capacity. New arrivals are routed to docks with low utilization first. In the case of a tie, the algorithm selects randomly between the charging docks it is indifferent between. An advantage of the lowest-utilization-first approach over other sorting methods (such as lowest-occupancy-first) is that it considers the different charging capacities of AC vs. DC docks in the assignment process.
- **Lowest-laxity-to-highest-capacity matching (LLHC)**: This strategy not only considers the state of individual charging docks but also that of the vehicle that is to be assigned. Specifically, it sorts arriving vehicles into baskets of low, medium, and high laxity (using bins obtained from historical data). Low-laxity vehicles are then matched with charging docks that have a high free capacity and vice versa, thus implicitly provisioning for future arrivals.

[4] This data is available via local transmission system operators (TSOs) and generally comes in 15 min intervals.

[5] C/C/C models assume constant battery charge/discharge efficiencies, constant energy content upper and lower bounds, and constant voltage (Kazhamiaka et al., 2019).

[6] For AC charging, the AC-DC inverter is integrated with the vehicle charger unit, whereas for DC charging the inverter sits inside the charging dock.

Vehicle charging algorithms. Vehicle-level charging schedules are re-computed in an online manner every five minutes allowing for updating of pre-computed schedules as new information becomes available. We implement a selection of sorting-based algorithms and optimization-based approaches to periodically determine the charge rate per connector k, vehicle j, and time t $\psi_{k,j,h,t}$:

- **First-come-first-served (FCFS)**: Charging requests are served on a first-come-first-served basis at full charging dock capacity until the available power capacity (on-site generation, storage, and grid) is exhausted. This algorithm is largely consistent with standard off-the-shelve load management tools available in the market today.
- **Least-laxity-first (LLF)**: Equivalent to a first-come-first-served algorithm but using a least-laxity-first priority rule meaning that least flexible vehicles are charged first. The algorithm, therefore, explicitly considers the current state of a vehicle in the charging decision.
- **Optimal**: Optimal operations uses mathematical optimization to periodically (re-)compute cost-optimal charging schedules that satisfy charging demand for the planning period δt in scope. We implement a standard cost-optimal charging framework (e.g., Ferguson et al., 2018). In our simulations, we plan $\delta t = 12$ periods ahead. We also implement a smoothing constraint by limiting the maximum charging ramp rate and considering the parallel use of docks. Since future arrivals and their preference vectors are unknown at the time of planning, we use a carefully tuned safety margin to be able to accommodate these in future periods. Further details are provided in A.1.

Electricity storage operational algorithms. The third and final system component requiring active operational management is the on-site energy storage system. We implement a heuristic approach that has been demonstrated to perform well under real-world conditions (Gust et al., 2021). We update battery-specific decision variables $\rho_{s,t}^{Charge}$ and $\rho_{s,t}^{Discharge}$ every five minutes[7]:

- **Temporal arbitrage (TA)**: The algorithms exploit the structure of the underlying time-of-use electricity tariff. During off-peak hours the battery is charged at a constant rate until the upper bound of the allowable energy content is reached. During on-peak periods, the battery is discharged at a constant rate to reduce the amount of electricity purchased at the higher on-peak price.

3.2.3. EVCH preference data

We populate the simulation's base architecture consisting of the above-described asset models and operational policies with an EVCH-specific high-resolution model of parking and charging preferences. EVCH user preferences (of an individual j) are described by the three-dimensional vector $v_j = (A_j, \delta_j, e_j^d)$ where A_j is the time of arrival, δ_j the duration of stay and e_j^d the requested energy. δ_j and e_j^d define what is referred to as laxity ($lax_j = \delta_j - \frac{e_j^d}{\kappa^i}$) (Lee et al., 2019). $lax_j = 0$ means that a vehicle j needs to charge at the maximum available rate κ^i for the entirety of its stay, while higher laxity values indicate more room for active charging management. Time of entry A_j determines the earliest planning period by which a certain charging event needs to be initiated. We start by building a model of current archetypical parking patterns. We do this in order to understand what typical parker types exist and how the user base composition can vary across facilities. A taxonomy of parker types can also be useful for building synthetic user population datasets based on assumed parker type shares wherever real-world data is not available. We leverage a unique, large-scale transaction-level parking dataset that was provided by a major European real-estate investor and includes transactions from seven large-scale parking garages.[8] We use a full year of data to capture daily, weekly and yearly seasonality. 2019 is chosen as a reference year to filter out pandemic-related effects. In total, our data comprises 3.84M parking events. We cluster parking events j based on A_j and δ_j, the two core parameters of interest at this modeling stage. All details related to data pre-processing, selection of clustering algorithms, and robustness tests are provided in e-companion B. In Table 2 we summarize our results. The largest proportion of parking events in our dataset is made up of three short-term parker types (*Morning Short*, *Afternoon Short* and *Evening Short*). These users enter a parking lot in the morning, afternoon, or evening respectively and typically stay for periods of 1–2 h. We also observe a *Business* cluster, which comprises parking events that commence in the early morning (7:26 am on average) and last for an average of 8 h. Two additional segments comprise longer-term parking events. These are *Overnight* parkers, which enter the parking lot in the late afternoon and stay until the next morning (typically 15.8 h on average), and *Long-term* parkers that stay for periods longer than 24 h on average.

We then look at the distribution of parker types across the different facilities in our dataset. Three archetypical facilities can be identified: The first facility type is a typical workplace facility that caters mostly to Business parkers. The second facility type is a destination facility. Apart from a small proportion of business users, such facilities mostly host short-term parkers. Finally, we also observe facilities with less conclusive usage patterns experiencing strong demand from all segments. We term these mixed-use facility.[9] Typical occupancy profiles for each of the three facility types are shown in Fig. 3.

Finally, we focus on the third required preference input variable: the requested energy per vehicle e_j^d. We employ a recently published real-world dataset by Lee et al. (2019) containing >25,000 charging transactions for the year 2019. Per each charging transaction the full preference vector $v_j = (A_j, \delta_j, e_j^d)$ is available. We blend the charging data (which only contains served sessions

[7] Note that battery decisions are made after the previously described charging decisions are made and that the available battery capacity at the start of each planning period is available for EV charging.

[8] Each row in this dataset represents a single parking event j with corresponding arrival and departure preference information. For privacy reasons, individual users cannot be identified.

[9] The example shown in Fig. 3 (right panel) is a large-scale inner-city parking facility that caters to workers, visitors, and residents.

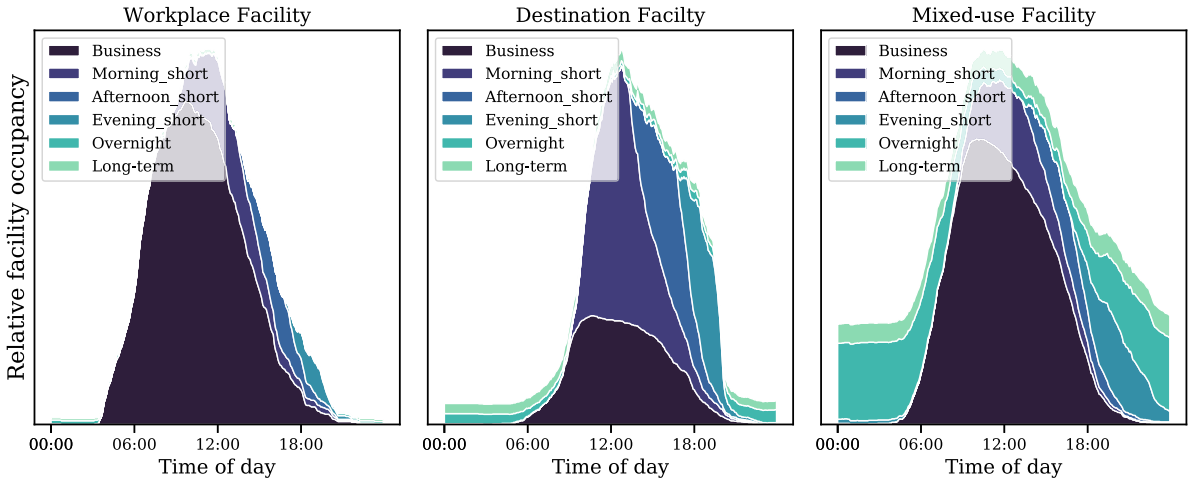

Fig. 3. Occupancy profiles per parker type (three archetypical parking facilities).

Table 2
Preference characteristics per parker type.

k	Name	Cluster size		Characteristics (avg & std. in parentheses)		
		N	share	A_j	δ_j	lax_j[a]
1	*Business*	671,384	17.47%	7:26 am (1.43 h)	7.92 h (2.73 h)	6.29 h (2.84 h)
2	*Morning Short*	1,279,646	33.30%	11:10 am (1.33 h)	2.12 h (1.90 h)	1.31 h (1.75 h)
3	*Afternoon Short*	985,710	25.65%	3:03 pm (1.00 h)	1.73 h (1.35 h)	1.08 h (1.30 h)
4	*Evening Short*	744,753	19.38%	6:17 pm (1.84 h)	1.47 h (1.30 h)	1.11 h (1.35 h)
5	*Overnight*	129,273	3.36%	5:22 pm (4.01 h)	15.84 h (4.02 h)	12.65 h (4.40 h)
6	*Long-term*	32,241	0.84%	2:28 pm (4.94 h)	37.04 h (6.70 h)	35.80 h (6.98 h)

[a] Assuming 22 kW max. charge rate.

that are constrained by the available infrastructure) with our parking dataset (which contains all parking requests per facility) using techniques from collaborative filtering. We train a prediction model on the labeled (Lee et al., 2019) dataset and use the resulting model to predict charging demand in the parking dataset. We obtain an exponentially distributed charging demand across the entire population of EVs with an average demand of 26.46 kWh ($\sigma = 17.20$ kWh) per parking session.[10] The distributional shape of charging demand is consistent with the one seen in other empirical EV charging settings (e.g., Ferguson et al., 2018). Crucially, however, Table 2 highlights important implications for charge management resulting from the different compositions of parker types in a facility. As can be seen, the average laxity varies significantly across parker types. Thus, parking facilities with a high proportion of high-laxity parkers (e.g., Business, Long-term) benefit from considerably higher flexibility characteristics with higher scope for optimization through intelligent charge management and parallel charging at lower rates.

3.2.4. Combining asset models, operational algorithms and preferences into an EVCH Digital Twin

Combining asset models, operational algorithms, and charging demand preferences yields a high-resolution DT of the envisioned EVCH system. Figs. 4 and 5 highlight the internal mechanics of the DT environment.[11] Fig. 4 represents the demand side and visualizes load curves for the various load sinks in the EVCH (EV charging, building baseload, battery storage charging) over the simulation horizon. Building load curves follow a recurring daily pattern ramping up during the day and down again during the night. Loads are slightly lower on the weekend (especially on Sunday). Note also the battery storage load, which reflects the temporal arbitrage strategy.

The power requested by the above-described load sinks is supplied by the grid, the on-site generation unit (PV), or the on-site electricity storage. The behavior of the supply side is shown in Fig. 5. Note how the exact supply mix heavily depends on the time of day (e.g., no PV generation after sunset, battery disc) and even weather conditions (note the considerably higher PV output in the middle of the week).

[10] We also apply some limited post-processing by limiting e_j^d to a realistic maximum bounded by the typical size of batteries (100 kWh) and feasible energy transfer over the duration of stay assuming 50 kW maximum charge rate.

[11] Shown here for a random week in a mixed-use facility using Lowest-laxity-to-highest-capacity routing, optimal charging and temporal arbitrage storage operations.

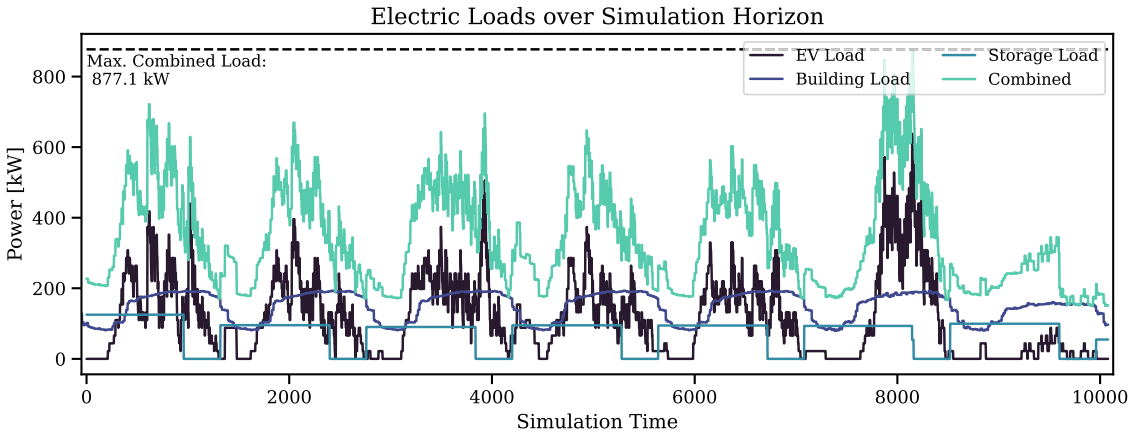

Fig. 4. Power supply by load sink over one-week simulation horizon (Monday to Sunday, Mixed-use facility).

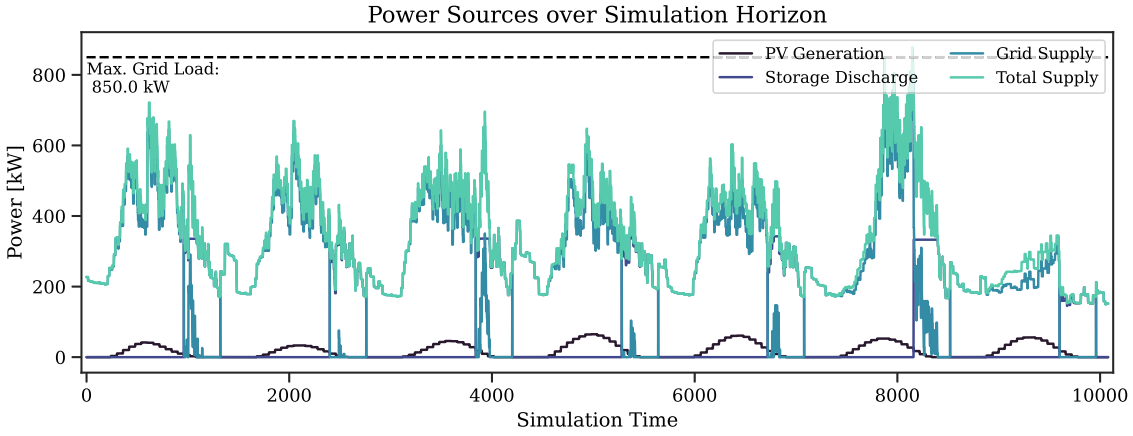

Fig. 5. Power supply by load source over one-week simulation horizon (Monday to Sunday, Mixed-use facility).

3.3. Setting up the SAC RL framework

Defining the design objective and the DT environment now allows us to address the system design challenge over multiple stages in an effort to obtain a (near-optimal[12]) solution.

For the case at hand and given that planning decisions are made over multiple stages h in planning horizon $\mathcal{H}$, the problem can be framed as a stochastic sequential decision-making problem. This decision process can be cast as a Markov Decision Process (MDP). An episodic task emerges, which starts at the beginning of the planning horizon ($h = 0$) and runs through the last investment stage $h = |\mathcal{H}|$, with the epochs being the individual decision stages (e.g., beginning of each year).

The MDP is formulated as a tuple $(S, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$, where the elements represent the state space, action space, transition probability, reward function, and the discount factor respectively. The state $s = (h, s^{chargers}, s^{grid}, s^{PV}, s^{storage})$ is defined as a vector of the time (i.e., the planning stage h) and the infrastructure which is accumulated over all previous stages. Actions are described by the vector $a = (a^{chargers}, a^{grid}, a^{PV}, a^{storage})$ and comprise planning decisions, such as the number of fast/slow docks with a specified number of plugs, the number of transformers, the PV capacity and storage to be installed. In line with the planning objective defined in Step ❶, we define the reward for moving from state s_h to s_{h+1} as $r = r(s, a) = -(C^{\Phi} + C^{\Omega} + C^{\Psi})$, which includes the investment cost, the operational costs, and C^{Ψ} which represents a penalty related to unserved charging demand. Note that the reward is the negative value of costs. Despite a deterministic state transfer function, exact reward functions are stochastic and unknown. In other words, given a decision, the next state is known, while the expected reward is unknown unless the state is evaluated in the EVCH DT environment. Therefore, we employ model-free reinforcement learning to find an optimal EVCH configuration policy π^*. The facility investment policy $\pi : S \times \mathcal{A} \rightarrow [0, 1]$ maps the state of the environment to an investment decision for each planning time

[12] Note that global optimality cannot be guaranteed for most learning or search methods.

step. Note that the output of the policy is standardized for all action components to be between 0 and 1. Furthermore, each action component is individually scaled based on the associated lower and upper bounds.

To find the optimal policy, two classes of RL algorithms have been proposed in the literature (Sutton and Barto, 2018): (1) value-based algorithms which learn the state/action-state values by interactions and shape the optimal policy using the learned values, and (2) policy-based algorithms which directly evaluate and improve the current policy until converging to near-optimal solutions.

As mentioned, value-based models such as Q-learning can run into issues of tractability in large state–action space environments like EVCH sizing. Therefore, we opt for a soft actor-critic (SAC) model which combines value-based and policy-based concepts. SAC is an off-policy actor-critic deep RL algorithm that works based on the maximum entropy learning framework (see Appendix C.1 for the differences between traditional actor-critic and SAC). In other words, the actor of SAC aims to learn a policy $\pi(a_h|s_h)$ that maximizes expected reward (negative values of costs) while also maximizing entropy to improve the exploration which is vital for large-scale problems. Therefore, the objective of learning the policy is defined as follows:

$$J(\pi) = \sum_{h=0}^{\mathcal{H}} \mathbb{E}_{s_h, a_h \sim \rho_\pi}[r(s_h, a_h) + \alpha \mathcal{H}(\pi(.|s_h))] \tag{2}$$

Where ρ_π denotes the marginal of the trajectory distribution induced by policy $\pi(a_h|s_h)$. The temperature parameter α adjusts the importance of the entropy term against the reward, and thus controls the stochasticity of the optimal policy. A key strength of SAC, setting it apart from other RL algorithms, is its powerful exploration capability. First, the policy is inherently stochastic, adding randomness to action selection. Second, the algorithm includes an entropy term in its objective function, promoting the exploration of less-visited policies. By adjusting the weight of this entropy term, SAC effectively balances exploration and exploitation. Additionally, random noise is introduced to the actions to prevent the model from getting stuck in local optima. For a detailed explanation of the SAC algorithm, we refer to Haarnoja et al. (2018).

To overcome the curse of dimensionality we use deep neural networks to represent both critic (value network) and actor (policy network). The value network evaluates the value of the current policy through interaction with the environment and the policy network makes decisions based on the current state of the system. Each network is fully connected and includes multiple (4) hidden layers with different number of nodes (256, 512, 512, 256) (See Appendix C for details). We train both networks from the agent's past experience using temporal-difference algorithms. This means that each experience in buffer contains one interaction with the environment, including state, action, immediate reward, and next state. To improve stability of the critic network, we also use a target network that gets updated less frequently and to increase the chance of more comprehensive exploration we add extra noise to the output of our policy network in the training phase.

In problems with multi-dimensional action spaces, using continuous-to-discrete mapping significantly enhances the scalability of the model compared to discrete action space models (e.g., Christodoulou, 2019), which must account for all possible action combinations. This scalability is crucial for addressing large-scale problems in real-world scenarios. We will illustrate this advantage by conducting a scalability analysis of our proposed model and comparing it with traditional discrete action space models. Our MDP contains discrete integer actions which may be better suited for discrete RL models at first glance. However, even state-of-art deep learning function approximation approaches do not easily scale to the number of action combinations (5^8 alternatives for each decision step) encountered in this problem. Alternatively, making use of integer relaxation, we can employ a continuous SAC model which is considerably more scalable than discrete (value-based) alternatives. Similar to Dulac-Arnold et al. (2015), our model first takes actions within a continuous space and then maps them to a discrete action set. We will define:

$$f_{\theta^\pi} : \mathcal{S} \rightarrow \mathbb{R}^D, f_{\theta^\pi}(\mathbf{s}) = \hat{\mathbf{a}} \tag{3}$$

f_{θ^π} is a function parameterized by θ^π (the policy network parameters), mapping from the state space $\mathcal{S}$ to the continuous action space $\mathbb{R}^D$, where D is the dimensionality of the action space. The output of this function ($\hat{\mathbf{a}}$) is likely not a valid set of actions for the environment as it will contain non-integer values that might violate the physical constraints of the environment. Therefore we define a mapping function as follows:

$$g : \mathbb{R}^D \rightarrow \mathcal{A} \tag{4}$$

$$g(\mathbf{s}, \hat{\mathbf{a}}) = argmin \sum_D (a_d - \hat{a}_d)^2 \tag{5}$$

$$a_d \in \mathcal{Z}^+ \cup \{0\} \ \forall d \in D \tag{6}$$

$$s_d + a_d \leq u_d \ \forall d \in D \tag{7}$$

g is a mapping function from a continuous space to a discrete space, constrained by the physical capacities of the environment.

Based on findings of Dulac-Arnold et al. (2015), although the architecture of our policy is not fully differentiable, we can nevertheless train our policy by following the policy gradient of f_{θ^π}. To do so we define a simpler policy $\pi_\theta = g \circ f_{\theta^\pi}$. In this initial case we can consider that the policy is f_{θ^π} and that the effects of g are a deterministic aspect of the environment. This allows us to adopt a standard policy gradient approach to train f_{θ^π} on its output $\hat{\mathbf{a}}$, effectively interpreting the effects of g as environmental dynamics. In addition to this, we include the physical restriction of each action component using constraint (7), whereby unfeasible actions are prevented. As an example of this, the number of installed PVs cannot exceed the area capacity of the EVCH. The addition of the continuous-to-discrete layer in the agent's policy is the primary modification compared to traditional SAC models. In order to ensure that the integer relaxation does not cause our model to converge to a local optimum, we benchmark the SAC model results

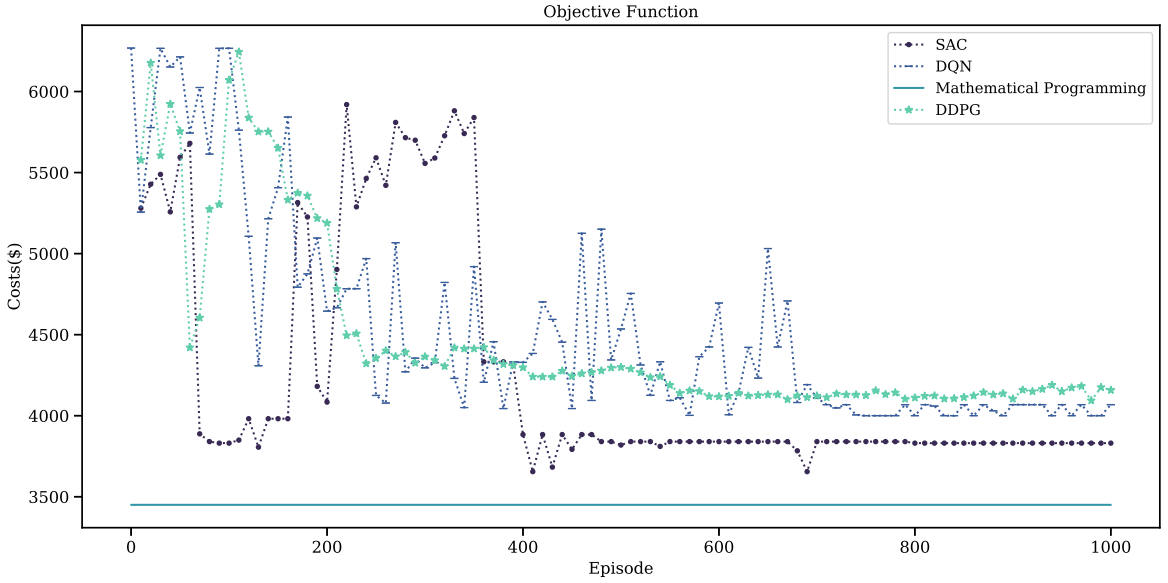

Fig. 6. Convergence and optimality characteristics of our proposed model as compared to DQN and mathematical programming - Values resampled to minimum out of 10 consecutive episodes.

with a deep Q-network agent which uses a discrete action space as well as with a perfect-information mathematical model. The DQN model does not rely on integer relaxation while the mathematical model represents the optimal solution under perfect information. The details of this comparison are provided in Section 4 and offer reassuring evidence that divergence to a local optimum is unlikely an issue in our experiments.

4. Performance evaluation and benchmark

We now evaluate our SAC-based RL model. We first present the convergence and optimality characteristics of our approach in comparison to traditional optimization methods, as well as popular reinforcement learning techniques, including the value-based Deep Q-Network (DQN) and the actor-critic Deep Deterministic Policy Gradient (DDPG) algorithms. Subsequently, we analyze the scalability properties of our SAC-based reinforcement learning model. All experiments are performed on a research workstation with an AMD Ryzen Threadripper 3970X 32-Core processor and 256 GB of RAM. To achieve tractability of the mathematical programming benchmark, we introduce several simplifications across the three modeling approaches to ensure a fair comparison. First, we limit the problem size to 200 parking spots. Second, facility operations are assumed to be optimal (optimal vehicle placement and charging) and stochasticity is neglected (i.e., assuming perfect foresight). Third, we reduce the operational detail of the EVCH simulation by adopting an hourly temporal resolution.

4.1. Evaluating convergence and optimality properties

In Fig. 6 we present convergence and optimality characteristics of our main model, the SAC-based RL framework, as compared against the optimal solution derived using mathematical programming. We also include two additional RL-based benchmark models in the comparison. First, we implement a value-based DQN reinforcement learner based on the algorithm presented in Van Hasselt et al. (2016) and used extensively in extant OM research. Second, we implement an alternative actor-critic method, the DDPG algorithm, based on Lillicrap (2015). This approach is more closely aligned with our proposed model, which utilizes the SAC algorithm. Details on all four models are presented in Appendix C.

In our experiments, SAC achieves a near-optimal solution in just 400 episodes, significantly outpacing DQN and DDPG, which converge after approximately 700 and 550 episodes, respectively. The optimality gap of the SAC model is significantly lower than that of the DQN and DDPG approaches (10% for SAC vs. 19% for DQN, and 22% for DDPG). We attribute these superior convergence and optimality gaps characteristics of SAC compared to DQN to the large state–action space of the given problem (8 discrete decision with 5 options each). Scalability to large state–action spaces is a well-known drawback of value-based approaches such as DQN which require more and more exploration to train larger and larger deep-Q networks (number of output nodes equals the number of actions that can be taken Dulac-Arnold et al., 2015). Indeed, this is confirmed in our experiments with larger/real-sized problem instances, where the DQN fails to converge in a reasonable amount of time (48 h). SAC can handle highly-dimensional problem spaces significantly better than DQN. This is due to the core “actor” component learning the policy function directly without the need to evaluate all possible actions per state. Instead, the actor returns actions directly in a continuous space, resulting in a smaller

Table 3
Evaluation of scalability and optimality.

EVCH facility size	20		50		100		200		500		1000	
	t	gap	t	gap	t	gap	t	gap	t	gap	t	gap
Mathematical program	4	0.00	248	0.00	842	0.00	5937	0.00	no solution		no solution	
DQN	932	0.05	1950	0.35	5420	0.14	18,582	0.19	no solution		no solution	
DDPG	343	0.05	1330	0.13	1780	0.12	2474	0.22	no solution		no solution	
SAC	**174**	**0.04**	**396**	**0.15**	**990**	**0.11**	**2115**	**0.10**	**15,919**	–	**42,714**	–

Notes: Problem size given by number of parking spots in facility, t indicates the solution/convergence time in seconds (excluding time needed for hyperparameter tuning); gap indicates the optimality gap; DQN: Deep Double Q-Networks RL model, DDPG: Deep Deterministic Policy Gradient RL model, SAC: Soft Actor-Critic RL model.

network. While DDPG leverages both actor and critic networks, it struggles with the trade-off between exploration and exploitation. The DDPG algorithm employs a deterministic policy and relies only on action noise added to the policy output for exploration. Our results demonstrate that in complex environments with high-dimensional action spaces, the SAC algorithm outperforms traditional actor-critic models like DDPG (Haarnoja et al., 2018).

4.2. Evaluating scalability to real-sized problems

We now run several additional experiments aimed at understanding scaling performance of the SAC-based planning approach and its performance against benchmark algorithms for smaller and larger problem sizes.

First, we explore whether SAC has advantages when it comes to scaling to practical problem sizes, i.e., EVCHs significantly larger than the previously explored 200 parking spots. To this end we iteratively increase the problem size from an initial facility size of just 20 parking spots to up to 1000 over the course of six experimental runs. 1000 parking spots is a size that is representative of a large parking lot (see Section 3.2.3). Results are shown in Table 3. Solution time as a function of EVCH facility size increases exponentially for the mathematical programming framework. The model requires just 4s to solve a 20 parking spot EVCH planning problem to optimality, but 5937 s to reach an optimal solution for a facility 10 times the size (200 parking spots). No convergence is achieved for problem sizes significantly larger than that (e.g., 500 parking spots and upward) meaning mathematical programming is not a practical option for many real-sized EVCH planning scenarios. Our SAC framework, on the other hand, exhibits significantly more favorable scaling and tractability properties. It achieves solution times 64% lower than mathematical programming for a 200 parking spot facility. It also scales to problem sizes of 1000 parking spots for which convergence is reached in just under 12 h. SAC converges very closely to the global optimum reaching optimality gaps between 4% to 15% across our experiments. Notably this is also reflected in the specific planning decision derived by SAC, which closely resemble optimal planning decisions. Please refer to Appendix E for a deeper analysis of individual planning decisions for all benchmark models.

Our results also highlight the scalability advantages of our proposed model (SAC) compared to value-based (DQN) and traditional actor-critic (DDPG) algorithms. While both DQN and DDPG perform close to optimal solutions for small facility sizes, their performance significantly deteriorates for sizes exceeding 200, and neither converges even after 1000 episodes. For DQN, convergence is slower even for small problems, and for larger facilities (e.g., 500 and 1000), the algorithm struggles to find good solutions due to the curse of dimensionality, as the action space becomes excessively large. Similarly, DDPG requires more time to converge compared to SAC, and the gap widens for large facilities. In such cases, DDPG fails to converge, primarily due to its reliance on insufficient exploration mechanisms, which is a limitation of traditional actor-critic models like DDPG (Colas et al., 2018).

As noted above, for reasons of comparability, these benchmark results are for an abstracted version of the EVCH sizing problem (perfect foresight, low temporal granularity and simplified operational detail). An additional benefit of adopting SAC versus mathematical optimization lies in the fact that operational and temporal detail do not increase the problem size and thus do not significantly impact solution time. This is because the modeling of operations is relegated to the DT simulation. Consequently, simulating real-sized EVCH systems in close to full operational detail, real-time (e.g., 1 min discretization) and over large sets of sensor data (e.g., months or even years) become possible. SAC-derived solutions can therefore be expected to generalize better to real-world stochastic conditions compared to the optimization-derived solutions. Note that stochastic and/or robust optimization approaches have not been explored in this work due to their significantly higher computational requirements compared to deterministic approaches which exacerbate scalability concerns.

5. Scenario analyses

Finally, we use our SAC model in conjunction with the DT simulator to run extensive sensitivity testing under close-to real-world conditions. Aspects of particular interest here are: user preference scenarios and operational policy choices and their impact on planning decisions. For illustrative purposes, we report configurations achieved in the final planning state (h=9) only, i.e., we neglect the scale-up pathway until that state is reached. Details on parameterization are provided in Appendix D.

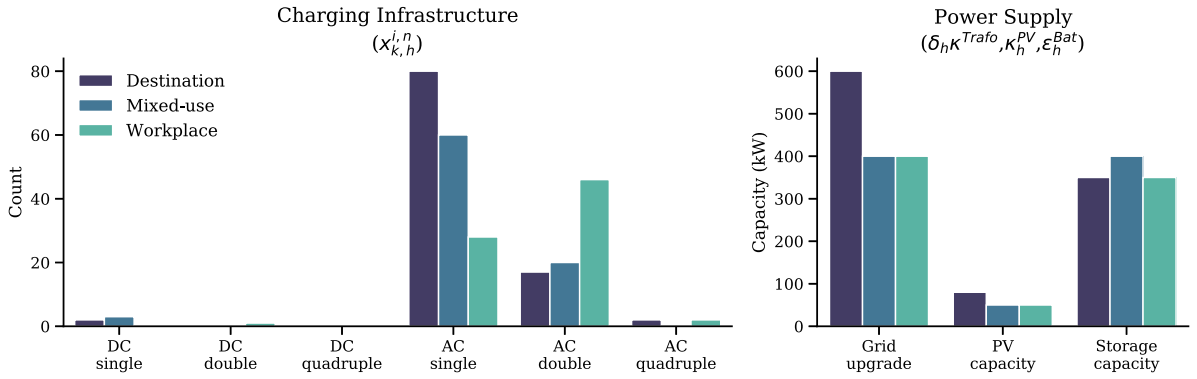

Fig. 7. RL-derived system configuration decisions for three archetypical EVCH facilities.

5.1. Impact of variations in user preferences

In Fig. 7 we compare SAC-derived infrastructure investment decisions for three archetypical EVCH facilities (Destination, Mixed-use, Workplace). The results reveal significant sensitivity of optimal physical layout decisions to user preferences. Recall that Destination facilities are primarily used for short-term parking (large proportion of *Morning Short*, *Afternoon Short* and *Evening Short* parkers), Mixed-use facilities exhibit a heterogeneous user pool, while Workplace facilities are primarily used by commuters with long stays (see Table 2 and Fig. 3). Consequently, average laxity characteristics vary considerably across facility types. For example, the average laxity of parkers in a Destination facility is considerably lower than that of a Workplace facility where users remain parked for prolonged periods. We see these preference differences reflected in the derived EVCH configurations across the three facilities. Specifically, for the facility with the lowest average laxity (Destination facility), the RL algorithm chooses to provision primarily single-connector AC docks along with a small number of single-connector fast chargers. A single-connector setup ensures that the full charging power is available at all times but comes at the risk of vehicles blocking an entire dock even after completing a charging cycle. The latter issue seems to be less problematic in a destination parking setting, where users do not stay long on average. At the other end, the infrastructure setup for a Workplace facility tends to favor multi-connector docks, particularly AC double-connector docks, thus taking advantage of the higher laxity of the underlying user population that affords longer charging cycles at lower rates. The third facility type (Mixed-use) falls somewhat in the middle between the previous two extremes. This is consistent with its laxity profile that lies between that of the Destination and Workplace facilities. It is noteworthy that DC charging plays a minor role in any scenario. EVCH charging use cases can mostly be satisfied at lower charging rates.

In terms of power supply infrastructure, the Destination facility requires an additional 200 kW transformer versus the other scenarios to achieve the target service level and satisfy low-laxity charging requests at higher charging rates. Some PV and battery storage is installed in all scenarios.

5.2. Impact of operational policy choices

Next, we leverage the flexibility characteristics of the DT approach by investigating the impact of different operational policies on the sizing decisions and the system's cost performance. As mentioned, such analyses would mean major model reformulation for optimization frameworks but are easily implemented in a DT-based model. In Fig. 8, we explore the impact of routing decisions for a Mixed-use facility, assuming FCFS charging operations. We find that the planning outcome is sensitive to the choice of routing strategy, as is the cost performance of the derived system. The more sophisticated routing strategy (LLHC) relies on more multi-connector docks and requires fewer alternative power sources (PV and storage) in an optimal setup compared to the same facility operated with LUF routing. Total system cost savings amount to 8.4%.

We perform a similar sensitivity analysis for the choice of charging strategy. Fig. 9 displays RL-derived infrastructure decisions and system cost performance for the same Mixed-use facility, assuming LLHC routing. We observe a largely similar picture here. Planning decisions are sensitive to the choice of charging strategies (e.g., more multi-dock chargers, more PV, and more battery with optimal strategy), as is cost performance (i.e., the best performance of the system with optimal charging with savings of 1.7 to 3.2% against the alternative strategies).

In sum, we show that the physical EVCH configuration (Π) is highly sensitive to Ω, the EVCH operational policy. In general, the more sophisticated the operational policies, the lower the total infrastructure requirements and the better the overall cost performance of the system. Thus, in order to obtain optimal planning decisions, operations managers need alignment on how they intend to operate the service system. Different operational strategies require different infrastructure layouts to achieve optimal performance and result in different total system costs.

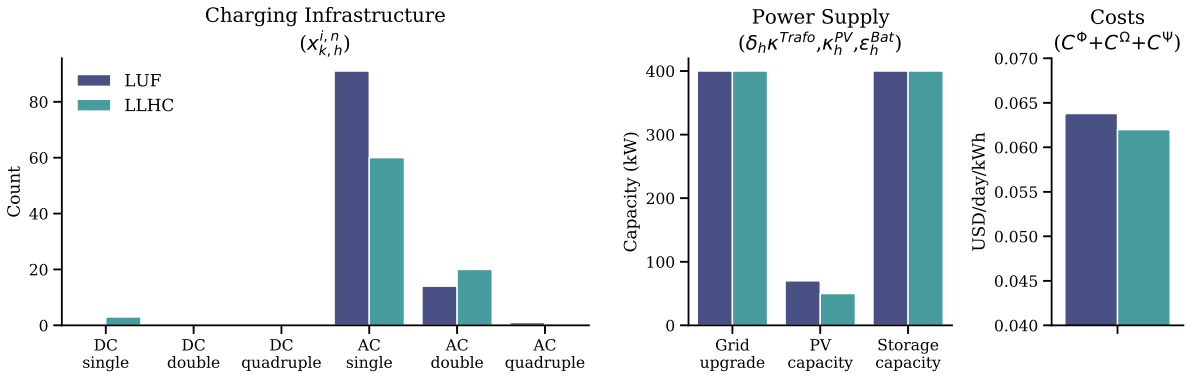

Fig. 8. RL-derived system configuration and performance against objective for different routing policies and a mixed-use facility.

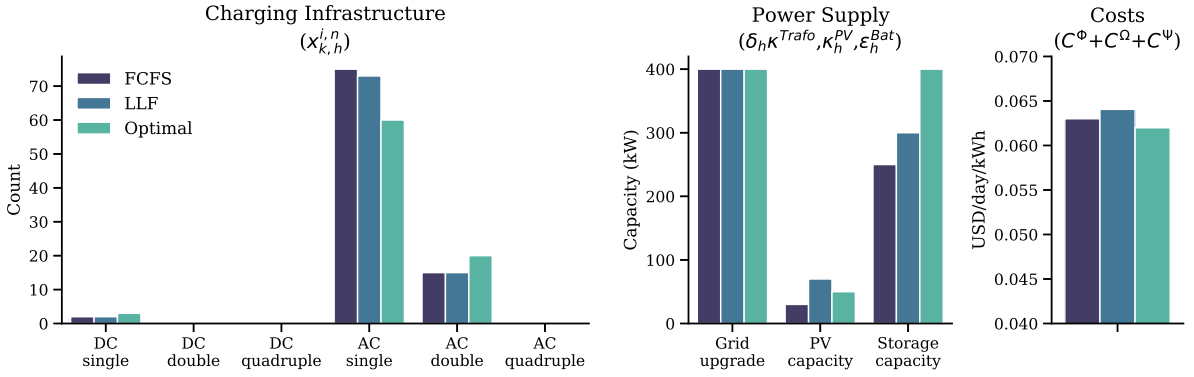

Fig. 9. RL-derived system configuration and performance against objective for different charging policies and a mixed-use facility.

6. Discussion

In this work we consider the problem of planning large-scale electric charging hubs (EVCHs). We define EVCHs as locally concentrated and centrally operated clusters of charging infrastructure that are typically integrated with on-site generation, storage and adjacent building infrastructure. Examples include workplace parking facilities, EV-enabled inner city parking garages or EV fleet depots. Planning these complex operational systems over a multi-year investment horizon represents a high-dimensional, dynamic and stochastic decision problem. Such planning problems have traditionally been approached by means of mathematical programming (e.g., Kazemi et al., 2016; Li et al., 2020). These frameworks are subject to computational challenges (e.g., NP-hardness) that can limit scalability to practical system sizes. As a result, simplifying assumptions related to, for example, temporal granularity, operational detail, system size, decision horizon or stochasticity are required to achieve tractability. This can come at the expense of generalizability to real-world conditions and does not take advantage of the wealth of granular operational data that has become increasingly abundant (Choi et al., 2022).

We develop and evaluate an alternative data-driven solution approach to the EVCH planning challenge, thus responding to calls from the scientific community and real-world sectors to develop methods that make use of and incorporate granular operational and preference data into OM frameworks (Qi and Shen, 2018; Cohen, 2018; Choi et al., 2022; Ketter et al., 2023).

The proposed solution – the core contribution of this work – leverages modern reinforcement learning (RL) (specifically soft actor-critic (SAC) RL) in combination with fine-grained data-driven simulation, also referred to in this work as Digital Twin (DT). SAC, a policy-based RL method, is better suited for the significant size of the action combinations in EVCH planning (e.g., wide variety of asset classes such as different EV charger types, PV, on-site battery, transformers, etc. each with large set of discrete options over multiple investment periods) compared to value-based deep learning function approximations. This is primarily due to the continuous nature of the action space in a SAC model. To adapt the continuous SAC model to the discrete action space of EVCH planning, our model first takes actions within a continuous space and then maps them to a discrete action set. We show that, for the case of EVCH, the proposed SAC-based model delivers on the key theoretical and practical benefits of RL: (1) scalability, (2) incorporation of operational detail and large-scale stochastic preference data, and (3) modeling flexibility. We also demonstrate that concerns around the lack of optimality guarantee are largely unfounded with our model converging closely to the global optimum across all our experiments despite the integer relaxation adopted in our approach. We provide further details on these key results in the following.

In terms of scalability, we demonstrate experimentally that using soft actor-critic RL in combination with a data-driven simulation environment is scalable to real-world EVCH system sizes of 1000 parking spots for which the model converges in under 12 h. Optimization methods (similar to the one proposed in Li et al. (2020)) only scale to EVCH facility sizes of approx. 200 parking spots and fail to converge (within the 48 h time constraint) for larger problems. An important caveat is the need for significant modeling simplifications, such as coarser temporal discretization, less operational detail and deterministic realization of normally uncertain parameters (e.g., arrival times, charging demand) to achieve tractability for these problem sizes with mathematical programming. Scalability (and optimality gap performance) of our SAC-based method also compares very favorably against alternative RL approaches, specifically the popular value-based DQN approach. For the 200 lot benchmark problem SAC converges within 400 episodes versus 700 episodes for DQN and achieves an optimality gap of just 10% versus 19% for DQN.

Another theoretical benefit of RL that we are able to exploit for this work on EVCHs and that distinguishes our method from extant optimization-based planning frameworks is the fact that RL scales almost independently of operational detail and data set size, due to the framework's reliance on a simulation rather than a mathematical model to capture system dynamics. This allows us to leverage data on preferences, operations and asset characteristics in great detail (e.g., 1 min intervals) and over long periods of operational data (e.g., months). For example, we use real-time parking and charging data to develop a novel taxonomy of parker types along and their charging preferences. We show that parking events can be classified into one of six categories (e.g., business parkers, overnight parkers, etc.) and that archetypical facility types (e.g., a workplace facility) exhibit very distinct parker population patterns. These data-driven and very granular preference models power our EVCH simulation and allow us to align our simulation with real-world conditions as much as possible. We posit that this will result in better performance of the target system under real-world conditions. Indeed, using high-detailed simulation environments both in terms of temporal granularity and preference granularity provides significant and quantifiable value. Our experiments reveal better cost performance of the derived planning decision vs. models where we either use coarser time periods (approximately 20% cost increase vs. the benchmark case as we increase modeling granularity to 2 h) or where we use distributional assumptions of preferences instead of real-world sensor data (significant 45% drop in service level compared to the benchmark case). While traditional stochastic or robust optimization approaches can be specified to account for uncertainty in future realizations of parameters, this comes at the cost of larger models and associated performance penalties rendering these alternatives impractical for the large-scale multi-stage stochastic EVCH planning problem.

As an added benefits of relying on a simulator rather than a mathematical model of the physical EVCH environment, the proposed method is extremely flexible. The key benefit of this modeling flexibility is the ability to conduct extensive scenario analyses regarding user preferences, operational policy choices, asset configurations and cost assumptions that have practical use for operations managers looking to make data-driven EVCH planning decisions. This can be achieved without requiring extensive model reformulation. We explore how infrastructure requirements change as asset operations become more sophisticated highlighting the value of such operational policies (see Section 5). We are also able to model complex interactions between a wide variety of different loads (building, EV charging, battery charging) and power sources (grid supply, PB, on-site battery storage), which sets this work apart from extant EVCH research (e.g., Kazemi et al., 2016; Li et al., 2020; Babic et al., 2022). These scenario analyses yield several interesting findings that have practical implications for EVCH operations management. For example, we find that integrating generation (PV) and storage (on-site) assets into the EV charging hub is beneficial in most cases, but varies by facility type and the adopted charging and routing policies. We also demonstrate that the use of multi-server chargers is cost effective and can improve EVCH economics, especially if active vehicle routing and smart charging strategies are adopted and if the user population has high average laxity as would be the case in a typical workplace facility. Another interesting finding is that, in many scenarios, no significant investment in DC fast charging is required to achieve the desired service level. Opportunities to build on, expand and tailor these scenario analyses while leveraging our RL approach abound and we leave them for future work.

There are also several limitations of RL which we have explored extensively in this work and which we lay out here.

First, RL does not guarantee optimality (Sutton and Barto, 2018). This concern may be further exacerbated by the need for integer relaxation of several discrete decision variables in the proposed SAC framework. To provide evidence to the contrary, we run extensive simulation experiments on different EVCH system sizes to explore how closely to the global optimum our solution converges. We show that the solutions obtained in these experiments can be considered near-optimal (gaps between 4% and 15%).

It should also be considered that despite the highly detailed nature of the DT simulation, it is still an abstraction of reality that is subject to several limitations and simplifications. For example, we consider routing and charging decisions separately instead of jointly. We also do not allow for vehicle-to-grid operations and consider loads of the attached building loads to be exogenous, among other simplifications. Such limitations represent exciting avenues for follow-up work, and we leave them for future research.

Finally, our method is data hungry, meaning that it requires large amounts of granular operational data. We argue, that with the emergence of inexpensive sensor technology, ubiquitous computing, and mobile connectivity such data tends to be increasingly available (Choi et al., 2022). While the above-mentioned benefits of RL should warrant these data and implementation these costs there is an additional argument in favor of the proposed method: as opposed to traditional OM planning models, the simulation environment (DT) used as part of the RL framework is not single-use. Indeed, the DT can be readily bridged-over into the use phase of the EVCH by simply replacing historical sensor data streams with real-time data flows (Boschert and Rosen, 2016). In the system's use phase, the DT then affords real-time system monitoring and optimization. We aim to exploit this multi-use characteristic in future work by leveraging the developed DT simulation in the development of novel high-performing learning algorithms for real-world EVCH operations.

In sum, our framework offers a novel practical method for OM practitioners to incorporate data-driven, high-fidelity simulators (i.e., DTs) combined with state-of-the-art reinforcement learning methods in the design phase of large-scale EVCH systems. The method yields near optimal planning solutions, scalability to real-world systems, the ability to incorporate and account for large-scale stochastic user and systems behavior as well as modeling flexibility.

CRediT authorship contribution statement

Karsten Schroer: Writing – original draft, Methodology, Data curation, Conceptualization. **Ramin Ahadi:** Writing – original draft, Methodology, Data curation, Conceptualization. **Wolfgang Ketter:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Conceptualization. **Thomas Y. Lee:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Operational modeling and algorithms

A.1. Optimal charging approach

Our charging model reconsiders the charging rate for connected vehicles at the beginning of each planning interval, thus following an online optimization paradigm.

The objective is to minimize the energy costs while meeting the charging demand of all connected vehicles (up to a predefined service level) over the look ahead planning window $\mathcal{T}^{\Omega}$. As we consider an uncertain case where the perfect information of upcoming vehicles is not available to the decision model, we make this conservative assumption to ensure that the service level is fulfilled.

First, the planning horizon must be chosen carefully due to its significant effect on model performance. In our simulations, we limit it to 6 h, which, given a planning interval Δ^{Ω} of 15 min, breaks down to 24 decision steps (i.e., charging rates are re-computed every 15 min of simulation time). We term the set of decision steps $\mathcal{T}^{\Omega}$

Second, we consider a flexibility margin μ_t to accommodate future, yet unknown demand. Although this model outputs a vector of charging rates for each vehicle, we only use the first charging rate and reconsider decisions in the next charging time step based on the updated system state including the new arrived vehicles.

$$Min_{\Xi} \sum_{t \in \mathcal{T}^{\Omega}} T_t^e e_t^{Grid} + T^p p^* \tag{8}$$

The grid energy consumption e_t^{Grid} accounts for the charging of vehicles (variables) as well as the storage (dis)charging, PV generation and building loads (storage rate is given as parameter before charging management). Constraint Eq. (10) guarantees that the grid energy consumption does not exceed the grid capacity minus the safety threshold we consider for the following time steps. We compute the induced peak in Eq. (11).

$$e_t^{Grid} = \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} \Delta_t (\psi_{k,j,t} + \beta_t^{Charge} - \beta_t^{Discharge} - f_t^{PV} \sum_{\tau=0} \kappa_{\tau}^{PV} + l_t) \ \forall t \in \mathcal{T}^{\Omega} \tag{9}$$

$$\frac{e_t^{Grid}}{\Delta_t} \leq p_t^{Grid} - \mu_t \ \forall t \in \mathcal{T}^{\Omega} \tag{10}$$

$$p^* \geq \frac{e_t^{Grid}}{\Delta_t} - l^* \ \forall t \in \mathcal{T}^{\Omega} \tag{11}$$

We also ensure that all vehicles receive at least η percentage of their charging demands (Eq. (12)). Constraint Eq. (13) ensures that vehicle can only charge when they are physically present in the EVCH. Finally, Constraint Eq. (14) restricts the parallel charging of vehicles that are connected to charging dock k to its charging capacity.

$$\sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}^{\Omega}} \psi_{k,j,t} \geq \eta e_j^D \ \forall j \in \mathcal{J} \tag{12}$$

$$0 \leq \psi_{k,j,t} \leq U_{j,t} M \ \forall k \in \mathcal{K}, j \in \mathcal{J}, t \in \mathcal{T}^{\Omega} \tag{13}$$

$$\sum_{j \in \mathcal{J}_k} \psi_{k,j,t} \leq \kappa^k \ \forall k \in \mathcal{K}, t \in \mathcal{T}^{\Omega} \tag{14}$$

Appendix B. Preference modeling

In this Section we provide details on the clustering routine and robustness test employed to identify parking archetypes. We cluster parking events j based on A_j and δ_j, the two core parameters of interest at this modeling stage. To account for the circular nature of arrival time A_j, which is not captured accurately by any distance-based clustering algorithm (for example, entries at 23:59 h and 0:00 h would be considered furthest apart despite their obvious proximity), we create two circular features $A_j^{sin} = \sin(2\pi(A_j/24))$ and $A_j^{cos} = \cos(2\pi(A_j/24))$. This yields the following vector of clustering variables $v_j^{clust} = (A_j^{sin}, A_j^{cos}, \delta_j)$, which we normalize.

Given the size of our dataset (3.84M observations) we limit our algorithm search to clustering algorithms that are sufficiently scalable. We run initial tests with three clustering algorithms: k-means++, a centroid-based algorithm, Gaussian Mixture Models (GMM) and BIRCH, a scalable density-based clustering algorithm. Overall, we find k-means++ to perform best in terms of runtime and stability. While GMM yields relatively similar results, BIRCH performs very poorly, yielding unstable and non-cohesive clusters

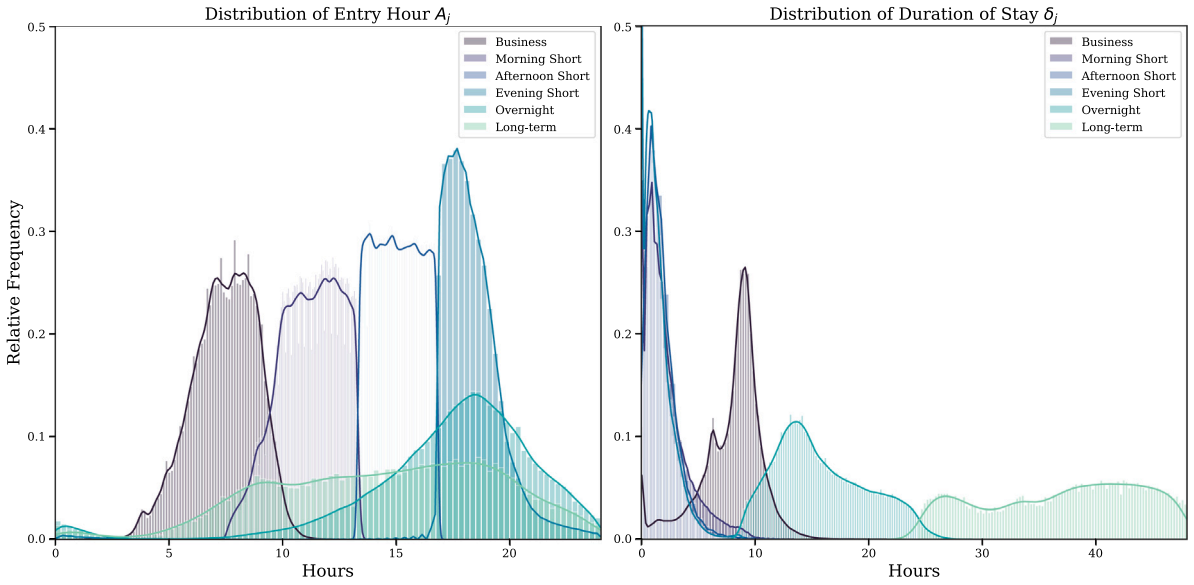

Fig. 10. Distribution of clustering variables per each cluster.

suggesting that relative density may not be a good identifier of clusters for the given dataset. We thus focus on fine tuning k-means++. A major challenge in the application of k-means++ is to select the number of centroids (clusters) k that are to be initialized and optimized for. To identify good candidate choices for k, We initially test integer values over an interval of reasonable values $[0, 20]$ and compute Calinski–Harabasz scores per each clustering outcome (Calinski and Harabasz, 1974). These analyses suggest $k = 5$ or $k = 6$ to be good choices. To validate and further narrow down our choice for k, we perform silhouette analyses for both candidate choices (Rousseeuw, 1987). We obtain the highest average silhouette coefficient $\bar{\mathcal{H}}$ for $k = 6$ ($\bar{\mathcal{H}} = 0.420$). Finally, taking $k = 6$ as the best performing choice across the above described internal validity measures, we conduct extensive cross-validation to assess cluster outcome robustness. We iteratively perform 2-1 splits of the data and re-run k-means++ on the larger dataset, then use the fitted algorithm to predict the labels of the smaller (test) dataset. We find our clustering results to be stable with observations in the test set having the same label 99.14% ($\sigma = 0.51\%$, 100 replications) of the time. We run an additional set of robustness analyses, this time focusing on the amount of preference data that is required to identify parker types reliably. This analysis draws on Griffin and Hauser (1993) who looked at the question of how many customer interviews were required for reliable insights. Clearly, there is a benefit to prospective EVCH planners if the need for data was smaller than the full one year period we have considered thus far. We run tests for 2 weeks, 4 weeks and 12 weeks of data per each facility and check the robustness of the clustering results as compared to the clusters obtained on the full one-year facility dataset. We find that high quality clustering results can be obtained with just three weeks of data (95.28%, $\sigma = 3.49\%$) accuracy vs. full one-year facility dataset. Beyond this threshold the value of additional data appears to diminish. At six weeks of data, for example, we obtain very similar accuracy (95.85% ($\sigma = 1.95\%$)), albeit slightly lower variance. In addition to internal validity and robustness of our clustering results we look at interpretability (or external validity). For this purpose, we presented our final clustering results (for $k = 6$) to a range of practitioners and discussed their implications. The clusters were deemed consistent with the domain experts' experience. In sum, we obtain six parker types that are supported both by internal criteria and real-world observation and can be readily identified with just three weeks of data. Fig. 10 shows the distributions of the two clustering variables per each final cluster.

Appendix C. Solution frameworks and models

In the following, we present the specifications of the model architectures developed and used in this research.

The common hyperparameters between the RL models are specified in Table 4. Model-specific parameters are explained separately. For each EVCH facility size, we tune the hyperparameters using a brute-force grid search, focusing on learning rate, batch size, training frequency, and hidden layers. The goal of the grid search is to find the best set of hyperparameters that converges to the highest objective function after training. Here we show the final hyperparameters for the main configuration used in the model benchmarking (200 parking spaces).

C.1. Main model: Soft Actor-Critic reinforcement learner

Actor-Critic and Soft Actor-Critic models are reinforcement learning algorithms designed to identify optimal policies for sequential decision-making problems, but they differ significantly in their approaches and objectives. Generally, Actor-Critic models

Table 4
Hyperparameter configuration for reinforcement learning algorithms.

Parameter	DQN	DDPG	SAC
Optimizer (learning rate)	Adam (0.001)	Adam (0.0001)	Adam (0.0001)
Loss function	MSE	MSE	MSE
Discount factor (γ)	0.99	0.99	0.99
Memory capacity	1,000,000	1,000,000	1,000,000
Steps prior to learning	1024	256	256
Training frequency	10	10	10
Batch size	64	256	256
Exploration strategy	Epsilon decay	Action noise	Entropy maximization
Target network update rate (τ)	0.01	0.01	0.05
Hidden layer activation function	ReLU	ReLU	ReLU
Number of hidden layers (nodes)	3 (256, 512, 256)	4 (256, 512, 512, 256)	4 (256, 512, 512, 256)

consist of two key components: the actor, which defines the policy function responsible for selecting actions, and the critic, which evaluates the actor's actions using a value function (Konda and Tsitsiklis, 1999). This framework aims to enhance policy performance by reducing variance in policy gradients through value-based feedback. SAC extends this model by including an entropy term in the objective function, encouraging the agent to balance exploration and exploitation by optimizing a trade-off between maximizing expected returns and policy entropy (Haarnoja et al., 2018). This objective improves training stability and sample efficiency, especially in environments with continuous or large action spaces. Additionally, SAC employs an off-policy approach, utilizing a replay buffer for more efficient data usage, whereas traditional Actor-Critic methods are often on-policy and require fresh data for updates (though not all Actor-Critic models are strictly on-policy). These characteristics make SAC particularly well-suited for complex, high-dimensional tasks requiring robust exploration and stability, such as the problem addressed in our study.

We utilize identical network architectures for both the actor and critic. We experiment with various hyperparameters for the actor and critic networks but found no significant impact on their performance, except for doubling the grid search size. The only contrast between the two is that we use a Tanh activation function in the final layer of the actor network, which yields a value between −1 and 1 for each sub-action and must be proportionally scaled based on the sub-action range. The SAC model utilizes Adam with a learning rate of 0.0001 to update the networks while implementing a batch size of 256. An automatic entropy tuning feature is also employed which is critical for SAC to minimize the impact of entropy in the objective function in order to achieve stable policies (Haarnoja et al., 2018). As suggested by existing literature (Haarnoja et al., 2018), a constant additional noise is applied during the training phase to enable the model to sufficiently explore the environment and mitigate the risk of local optima. The noise function follows a normal distribution with a mean of zero and variance of 0.05. We use Kaiming uniform initialization (also known as He initialization), which is the default weight initialization for a fully connected (feedforward) neural network (FNN) layer. This initialization is well suited for layers with ReLU or similar activation functions, as Kaiming initialization helps to maintain the variance of activations across layers. In addition, our model set the initial values of the biases to zero.

C.2. Benchmark RL model: DQN reinforcement learner

We use deep double Q-networks as the benchmark model in our study. For further details, please refer to Van Hasselt et al. (2016). The rationale behind Double Q-learning is to mitigate overestimation by breaking down the maximum operation in the target into action selection and evaluation. This results in the usage of two networks, one each for action selection and evaluation purposes. The action evaluation network is termed target network and undergoes lesser updates compared to the main network. To adjust the parameters of the target network, we adopt soft update, which is implemented as follows:

$$\theta' \leftarrow (1-\tau)\theta' + \tau\theta \tag{15}$$

Where θ' represents the target network parameters, θ represents the main network parameters, and τ is the soft update weight, which ranges from zero to one. The hyperparameter grid search, as shown in Table 1, indicates that DQN requires a higher learning rate, lower batch size, and smaller hidden layers compared to SAC. Our exploration strategy employs an epsilon-decay algorithm, where random actions are chosen with decreasing probability of epsilon during the training process. In our model, epsilon begins at 0.3 and decreases to 0.01 after 600 episodes, and remains unchanged thereafter.

C.3. Benchmark RL model: Deep deterministic policy gradients

Deep Deterministic Policy Gradients (DDPG) is a model-free, off-policy reinforcement learning algorithm designed for continuous action spaces (Lillicrap, 2015). It combines the strengths of both Q-learning and policy gradient methods. Similar to SAC, DDPG utilizes an actor-critic architecture, where the actor network learns a deterministic policy to map states to actions, and the critic network evaluates the Q-value of the state–action pairs. Inspired by Deep Q-Networks (DQN), DDPG uses a replay buffer to store experiences and sample mini-batches for training, ensuring decorrelated updates and improved stability. Additionally, we employ target networks for both the actor and critic to stabilize training by providing a slowly updated, consistent set of parameters. To encourage exploration in deterministic policies, we add noise to the actions during training. Although DDPG is designed for continuous action spaces, we adapt it to the integer action space of our problem using the same modifications applied to our proposed SAC model.

C.4. Upper bound benchmark: Mathematical programming model

A Mathematical Programming Model acts as upper bound in our benchmarks and is used to compute optimality gaps of RL-based models. We formulate the decision challenge as a feasibility problem which aims to satisfy all or a specified amount of total charging demand most resource efficiently while considering rate, space, and total capacity constraints. In doing so we expand on and adapt extant EVCH planning models (e.g., Li et al., 2020).

In line with the planning objective, we frame the problem as a cost minimization planning with the goal to jointly minimize the investment cost (C^{Φ}) and the operations cost (C^{Ω}) of the EVCH while ensuring a certain service level η_h^{Serv}. Formally, the objective can be expressed as follows:

$$Min_{\Xi}[(C^{\Phi}(x_{k,h}^{i,n}, \delta_h^{Trafo}, \kappa_h^{PV}, \epsilon_h^{Bat}) + C^{\Omega}(\omega_{k,j,h}, \psi_{k,j,h,t}, \beta_{h,t}^{Charge}, \beta_{h,t}^{Discharge})] \tag{16}$$

Both cost items are defined as follows. The investment cost (C^{Φ}) is the sum of the grid expansion cost (if any), the cost of charging infrastructure plus any installed PV and battery capacity over the full investment horizon $\mathcal{H}$. The operations cost (C^{Ω}) is defined as the total sum of electricity costs over the investment horizon, where costs are only incurred on the electricity retrieved from the grid with $e_{h,t}^{Grid}$. Formally:

$$C^{\Phi} = \sum_{h\in\mathcal{H}}[(c_h^T \delta_h^{Trafo} + \sum_{k\in\mathcal{K}} c_h^{i,n} x_{k,j,h}^i + c_h^{PV} p_h^{PV} + c_h^{Bat} \epsilon_h^{Bat})(1 + (|\mathcal{H}| - h)\mu^{Maint})] \tag{17}$$

$$C^{\Omega} = \sum_{h\in\mathcal{H}} \sum_{t\in\mathcal{T}} T_{h,t}^e e_{h,t}^{Grid} + T_h^p p_h^* \tag{18}$$

Note that $e_{h,t}^{Grid}$ is accounted for on the basis of a two-part tariff charging for both the use of electricity from the grid (excl. PV generation and possible battery discharge $\beta_{h,t}^{Discharge}$) and demand charges arising from the induced peak load attributable to EVCH operations. Demand charges T_h^p are designed to incentivize efficient utilization of the grid (Gust et al., 2021) and are typically based on the monthly peak load induced by the facility. We therefore define p^* as the excess of the expected base facility peak load l^* (excl. EVCH operations) for state h (Eq. (20)).

$$e_{h,t}^{Grid} = \sum_{j\in\mathcal{J}_h} \sum_{k\in\mathcal{K}} \Delta_t(\psi_{k,j,h,t} + \beta_{h,t}^{Charge} - \beta_{h,t}^{Discharge} - f_{h,t}^{PV} \sum_{\tau=0}^{h} \kappa_\tau^{PV} + l_{h,t}) \tag{19}$$

$$p_h^* \geq \frac{e_{h,t}^{Grid}}{\Delta_t} - l_h^* \;\; \forall h \in \mathcal{H}, t \in \mathcal{T} \tag{20}$$

The optimization is subject to additional operational and physical constraints.

$$\sum_{k\in\mathcal{K}} \sum_{t\in\mathcal{T}} \psi_{k,j,h,t} \geq \eta_h^{Serv} e_j^D \;\; \forall h \in \mathcal{H}, j \in \mathcal{J}_h \tag{21}$$

$$x_{k,h}^{i,n}, \omega_{k,j,h} \in \{0,1\} \;\; \forall k \in \mathcal{K}, h \in \mathcal{H}, j \in \mathcal{J}_h, i \in \mathcal{I}, i \in \mathcal{N} \tag{22}$$

$$\sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}} \sum_{n\in\mathcal{N}} \sum_{h\in\mathcal{H}} x_{k,h}^{i,n} n \leq L \tag{23}$$

$$\sum_{h\in\mathcal{H}} \sum_{i\in\mathcal{I}} \sum_{n\in\mathcal{N}} x_{k,h}^{i,n} \leq 1 \;\; \forall k \in \mathcal{K} \tag{24}$$

$$\sum_{j\in\mathcal{J}_h} \omega_{k,j,h} U_{j,h,t} \leq \sum_{\tau=0}^{h} \sum_{i\in\mathcal{I}} x_{k,\tau}^{i,n} n \;\; \forall k \in \mathcal{K}, h \in \mathcal{H}, t \in \mathcal{T} \tag{25}$$

$$\sum_{k\in\mathcal{K}} \omega_{k,j,h} \leq 1 \;\; \forall h \in \mathcal{H}, j \in \mathcal{J}_h \tag{26}$$

$$0 \leq \psi_{k,j,h,t} \leq \omega_{k,j,h} U_{j,h,t} M \;\; \forall k \in \mathcal{K}, h \in \mathcal{H}, j \in \mathcal{J}_h, t \in \mathcal{T} \tag{27}$$

$$\sum_{j\in\mathcal{J}_h} \psi_{k,j,h,t} \leq \sum_{\tau=0}^{h} \sum_{i\in\mathcal{I}} \sum_{n\in\mathcal{N}} x_{k,\tau}^{i,n} \kappa^i \;\; \forall k \in \mathcal{K}, h \in \mathcal{H}, t \in \mathcal{T} \tag{28}$$

$$\sum_{h\in\mathcal{H}} \kappa_h^{PV} \leq R \tag{29}$$

$$SoC_{h,t} = SoC_{h,t-1} + (\beta_{h,t-1}^{Charge} - \beta_{h,t-1}^{Discharge})\Delta_t \;\; \forall h \in \mathcal{H}, t \in \{1, 2, \ldots, \mathcal{T}\} \tag{30}$$

$$SoC_{h,0} = \sum_{\tau=0}^{h} SoC^{min} \epsilon_\tau^{Bat} \;\; \forall h \in \mathcal{H} \tag{31}$$

$$\beta_{h,t}^{Charge} \leq \beta_{h,t}^{Direction} \beta^{max} \;\; \forall h \in \mathcal{H}, t \in \mathcal{T} \tag{32}$$

$$\beta_{h,t}^{Discharge} \leq (1 - \beta_{h,t}^{Direction}) \beta^{max} \;\; \forall h \in \mathcal{H}, t \in \mathcal{T} \tag{33}$$

$$\beta_{h,t}^{Charge} \Delta_t \leq \sum_{\tau=0}^{h} SoC^{max} \epsilon_\tau^{Bat} - SoC_{h,t-1} \;\; \forall h \in \mathcal{H}, t \in \{1, 2, ..\mathcal{T}\} \tag{34}$$

$$\beta_{h,t}^{Discharge} \Delta_t \leq SoC_{h,t-1} \;\; \forall h \in \mathcal{H}, t \in \{1,2,..\mathcal{T}\} \tag{35}$$

$$e_{h,t}^{Grid} \leq \Delta_t(\kappa_0^{Grid} + \sum_{\tau=0}^{h} \delta_\tau^{Trafo} \kappa^{Trafo}) \;\; \forall h \in \mathcal{H}, t \in \mathcal{T} \tag{36}$$

First and foremost, service level is guaranteed in Eq. (21). Note that the summation is bounded by set $\mathcal{T}$, meaning that we consider the total supplied energy at the time of departure. This important constraint ensures that adequate infrastructure is provisioned despite the cost minimization objective.

EV charging infrastructure decisions and operations are controlled by means of decision variables $x_{k,h}^{i,n}$, $\omega_{k,j,h}$ (both binary indicators, see Eq. (22)) and $\psi_{k,j,h,t}$. First, the number of charging docks and associated connectors is restricted by the space constraints L of the facility (Eq. (23)). Similarly, Eq. (24) ensures that candidate points can only be equipped with chargers once and that this decision cannot be changed over the planning period, i.e., they cannot be removed once installed.

In terms of routing and charging operations, the model assigns vehicles to chargers upon arrival (one-off decision) and periodically adjust the charging power over the duration of their visit. Constraint Eq. (25) allocates vehicle j to spot k during stage h only if k is equipped with a charging dock and only if j is present in the EVCH (captured via $U_{j,h,t}$). Eq. (28) ensures that each vehicle connects to at most one charging dock. Constraint Eq. (27) guarantees that vehicle j receives non-negative energy (bounded by the maximum power of the specific dock in Eq. (28)) from charging dock k only if it is connected to k.

Battery and on-site generation constraints are set as follows. We assume PV generation to be non-controllable meaning no constraints are necessary to model their operations (in-feed is an exogenous parameter). We simply limit the maximum installable PV capacity $\sum_{h\in\mathcal{H}} \kappa^{PV}$ to the available on-site space (such as rooftop space) R (see Eq. (29)). Eq. (30) through (35) implement various battery-related constraints. Constraint Eq. (30) incrementally updates the battery state of charge $SoC_{h,t}$. Constraint Eq. (31) ensures that the battery SoC remains within a certain interval. Note that we neglect efficiency losses and assume battery depreciation to be independent of operations (Sharifi et al., 2020). We realize that these are simplifications, yet these are necessary to retain tractability of our model. We assume symmetric charge/discharge rate limits which are enforces through constraint Eqs. (32) and (33), where $\beta^{max} \geq 0$. These constraints also ensure that the battery cannot be charged and discharged at the same time.

Our model ensures that the EVCH's base load as well as EV and battery charging loads cannot exceed the total grid capacity (existing and extension) plus current PV generation, which is enforced by Eq. (36). Note that if the battery was discharging (negative β_t^{Bat}) this would increase the available capacity.

Appendix D. Experimental setup

In this Section we provide details on the experimental setup and parameterization of the Digital Twin (DT) simulation environment. Table 5 provides details on the key components of the DT environment and the digitalization approach adopted (real-world sensor data vs. simulation).

Note that we rely on real-world sensor data to model system dynamics wherever available and resort to simulations informed by research papers and/or asset specification sheets in all other cases (following Sierla et al., 2018). In addition, we impose several physical constraints inherent to the various components of the EVCH system. These are summarized in Table 6.

Table 7 summarizes the core investment-related parameters (costs and space constraints) used in the benchmark experiments over the investment horizon (10 states $s \in S$). Energy costs are based on real-world electricity tariffs from the same region in which the charging data were gathered (i.e., California). Table 8 gives an overview of the tariff structure that is used throughout all experiments.

Appendix E. Supplemental results

E.1. Investment decision by time period

In this Section we take a closer look at the dynamics of the investment decisions derived by the four decision frameworks implemented in this work (optimal, DQN, DDPG, SAC). This supplements the benchmark results shown in Section 4. It allows us to analyze the differences in planning decisions between the different planning algorithms in much more detail. Figs. 11 through 14, show the investment plans for charging infrastructure (left) and power supply infrastructure (right) derived via mathematical programming (optimal), DQN, DDPG and SAC, respectively. Note that in order to achieve comparability with the benchmark mathematical model, the experiments run in Section 4 do not consider on-site storage, hence storage is not being built in any planning period.

There are a few very interesting insights to be taken from this analysis. First, the planning decisions made by SAC are very close to the optimal decisions. Under both investment plans charging demand is primarily served through AC charging using 4 connectors per charger. Although the scale up path is slightly different for the supply side, both decision algorithms ultimately arrive at the same end state (100kw PV plus 200 kW of grid extension). This further underlines the robust and near-optimal performance of SAC. DDPG achieves the same supply decisions. However, it relies on significantly more charging infrastructure including a large amount of single-server AC chargers yielding suboptimal cost performance vs. the upper limit optimal investment plan and SAC.

DQN takes a notably different set of planning decisions compared to the other decision frameworks. This applies both to the number of charging docks that are being installed and to the investments in power supply. DQN installs the smallest number of

Table 5
Digital Twin (DT) components and associated datasets.

DT component	Type	Digitalization approach	Description	Source
Local Substation	Physical Asset	simulated	Transformation losses and physical limit	assumptions as detailed in Table 6
On-site Electricity Generation Assets (PV Panels)	Physical Asset	real-world sensor data	PV load factors (power output as percentage of installed capacity)	Open Power System Data provided by Neon Neue Energieökonomik and Technical University of Berlin and ETH Zürich and DIW Berlin (2024)
Electricity Storage Assets	Physical Asset	simulated	(dis-)charging efficiency curves, physical constraints (min/max state of charge)	Ghiassi-Farrokhfal et al. (2016); assumptions as detailed in Table 6
EV Charging Docks and Connectors	Physical Asset	simulated	AC-DC conversion losses, maximum charging capacity	assumptions as detailed in Table 6
Peripheral Building Electricity Consumption	Preference Pattern	real-world sensor data	Peak load in KW and consumption in kWh per 15-min interval	Unique real-world meter data provided by a major European real-estate investor; includes peak building loads and consumption at 15 min resolution across thirteen facilities with different usage profiles
Parking Demand	Preference Pattern	real-world sensor data	Vehicle-level time of arrival and duration of stay obtained from parking garage sensors	Unique transaction-level parking dataset provided by a major European real-estate investor; includes transactions from seven large-scale parking garages catering to different parking use cases (office building, destination parking, mixed-use)
Charging Demand	Preference Pattern	real-world sensor data	Requested energy per charging session in kWh	Real-world dataset by Lee et al. (2019) containing >25,000 charging transactions for the year 2019

Table 6
Operational constraints and parameters.

	Symbol	Value	Unit
Substation transformation efficiency	η^{Trafo}	98	%
PV DC-AC inversion efficiency	η^{Inv}	96	%
EV charging efficiency	η^{EV}	95	%
Energy storage charging/discharging efficiency	η^{Bat}	95	%
Energy storage minimum storage SoC	SoC^{Min}	5	%
Energy storage maximum storage SoC	SoC^{Max}	95	%
Energy storage maximum charging rate	κ^{Bat}	dependent on battery size; full cycle in 1h	kW
Standard size of a transformer	κ^{Trafo}	200	kW
Initial capacity of the grid connection	κ_0^{Grid}	250	kW

Table 7
Parameterization of benchmark experiments.

Parameter[a]	Unit	s_0	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	Source
Battery cost (c_s^{Bat})	USD/ kWh	575	507	441	391	352	321	295	273	255	239	Lazard, Bloomberg NEF
AC charger cost[b] ($c_{s,AC}^{EVSE}$)	USD/ unit	4500	4322	4151	3986	3828	3677	3531	3391	3257	3128	industry quotes
DC charger cost[c] ($c_{s,DC}^{EVSE}$)	USD/ unit	50 000	49 000	47 060	45 196	43 406	41 687	40 037	38 451	36 928	35 466	California Energy Commission
Connector cost AC	USD/ unit	250	250	250	250	250	250	250	250	250	250	assumption
Connector cost DC	USD/ unit	2500	2500	2500	2500	2500	2500	2500	2500	2500	2500	assumption
EV share[d] (σ_s^{EV})	%	5	8	12	18	27	37	42	49	56	65	Bloomberg NEF
Grid cost[e] (c_s^{Grid})	USD/ kW	250	276	304	335	369	407	449	495	546	602	industry quotes
PV cost (c_s^{PV})	USD/ kWp	2125	2041	1960	1882	1808	1736	1668	1601	1538	1477	Lazard, IEA
Max. PV capacity[f] (R)	kWp	100	100	100	100	100	100	100	100	100	100	–
Number of parking spots (S)	units	200	200	200	200	200	200	200	200	200	200	–

[a] All cost parameters include cost of installation and peripheral equipment (e.g., inverters for battery, PV and DC chargers).
[b] 22 kW, single connector.
[c] 50 kW, single connector.
[d] Assuming sales share equals penetration for given facility.
[e] Cost of transformer, cabling and contribution to upstream grid upgrades; assuming 5% yearly cost increase.
[f] Assuming 500 m^2 roof space and PV energy density of 0.2 kWp/m^2.

Table 8
Time-of-use tariff and demand charge for large-scale EV charging customers (> 500 kW).

	Summer (Jun - Sep)	Winter (all other months)
Super Off-Peak (8am-4pm)	0.08 USD/kWh	0.06 USD/kWh
On-Peak (4pm to 9pm)	0.23 USD/kWh	0.23 USD/kWh
Off-Peak (9pm-8am)	0.08 USD/kWh	0.08 USD/kWh
Demand Charge (monthly)	15.48 USD/kW	

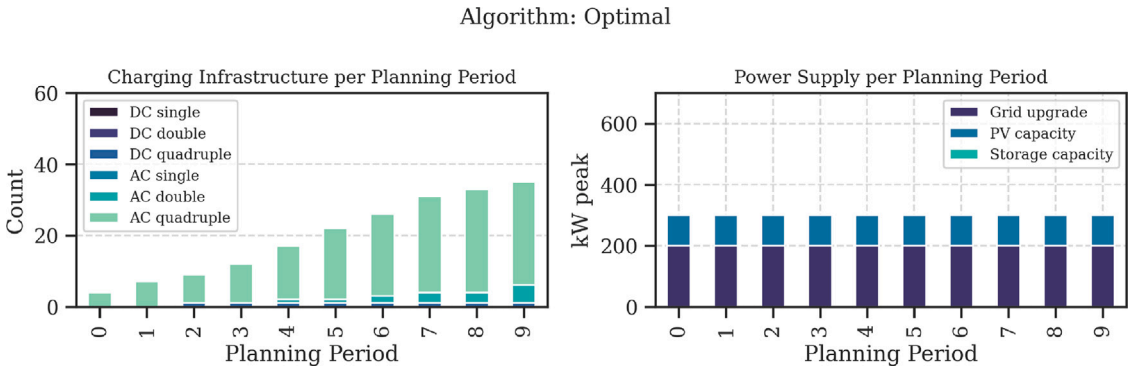

Fig. 11. Investment decision derived by the upper limit mathematical programming model.

docks and is the only framework to install DC chargers in significant numbers. Although this charging infrastructure may be able to serve the charging demand (primarily by providing higher average charge rates), this comes at the expense of significantly higher combined peak loads. As a result, the charging cluster configuration suggested by the DQN algorithm requires roughly twice the amount of grid updates to serve these loads compared to the other benchmark algorithms. This is suboptimal from a total cost perspective and highlights the value of multi-server charging docks and lower charging rates.

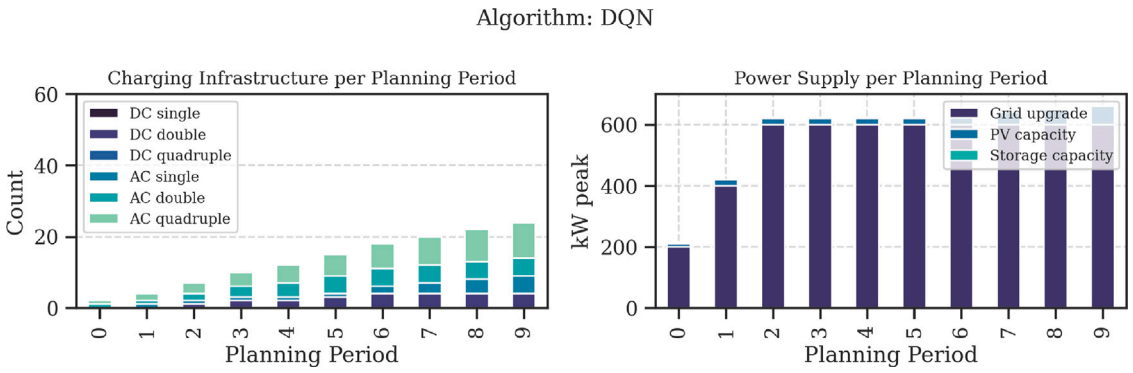

Fig. 12. Investment decision derived by DQN.

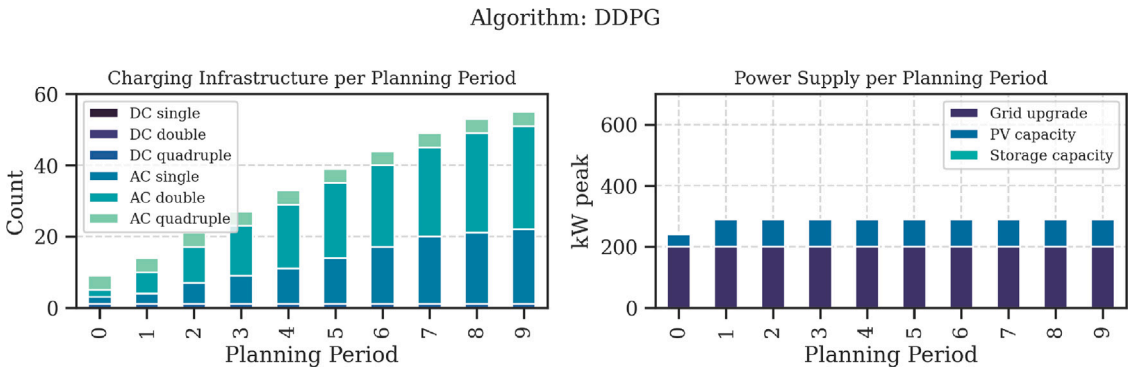

Fig. 13. Investment decision derived by DDPG.

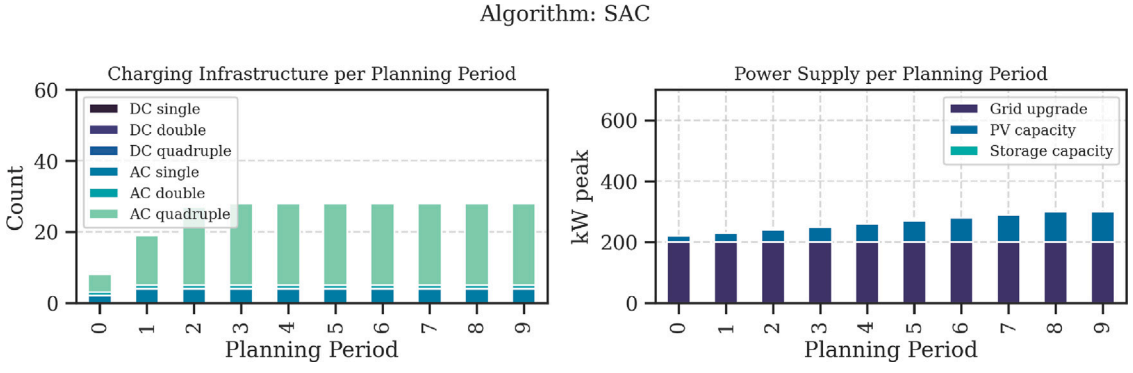

Fig. 14. Investment decision derived by SAC.

References

Ahadi, R., Ketter, W., Collins, J., Daina, N., 2023. Cooperative Learning for Smart Charging of Shared Autonomous Vehicle Fleets. Transp. Sci. 57 (3), 613–630. http://dx.doi.org/10.1287/trsc.2022.1187, http://pubsonline.informs.org/doi/10.1287/trsc.2022.1187. https://pubsonline.informs.org/doi/10.1287/trsc.2022.1187.

Anon, 2015. Handbook of Simulation Optimization. International Series in Operations Research & Management Science, vol. 216, Springer New York, New York, NY, http://dx.doi.org/10.1007/978-1-4939-1384-8,

Attaran, M., Celik, B.G., 2023. Digital twin: Benefits, use cases, challenges, and opportunities. Decis. Anal. J. 6, 100165. http://dx.doi.org/10.1016/j.dajour.2023.100165, URL https://www.sciencedirect.com/science/article/pii/S277266222300005X.

Babic, J., Carvalho, A., Ketter, W., Podobnik, V., 2022. A data-driven approach to managing electric vehicle charging infrastructure in parking lots. Transp. Res. Part D: Transp. Environ. 105, 103198. http://dx.doi.org/10.1016/j.trd.2022.103198.

Boschert, S., Rosen, R., 2016. Digital Twin—The Simulation Aspect. In: Mechatronic Futures. Springer International Publishing, Cham, pp. 59–74. http://dx.doi.org/10.1007/978-3-319-32156-1_5.

Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Comm. Statist. Theory Methods 3 (1), 1–27. http://dx.doi.org/10.1080/03610927408827101, URL http://www.tandfonline.com/doi/abs/10.1080/03610927408827101.

Choi, T.M., Kumar, S., Yue, X., Chan, H.L., 2022. Disruptive Technologies and Operations Management in the Industry 4.0 Era and Beyond. Prod. Oper. Manage. 31 (1), 9–31. http://dx.doi.org/10.1111/poms.13622.

Christodoulou, P., 2019. Soft actor-critic for discrete action settings. arXiv preprint arXiv:1910.07207.

Cimino, C., Negri, E., Fumagalli, L., 2019. Review of digital twin applications in manufacturing. Comput. Ind. 113, 103130. http://dx.doi.org/10.1016/j.compind.2019.103130.

Cohen, M.C., 2018. Big Data and Service Operations. Prod. Oper. Manage. 27 (9), 1709–1723. http://dx.doi.org/10.1111/poms.12832.

Colas, C., Sigaud, O., Oudeyer, P.Y., 2018. Gep-pg: Decoupling exploration and exploitation in deep reinforcement learning algorithms. In: International Conference on Machine Learning. PMLR, pp. 1039–1048.

Dulac-Arnold, G., Evans, R., van Hasselt, H., Sunehag, P., Lillicrap, T., Hunt, J., Mann, T., Weber, T., Degris, T., Coppin, B., 2015. Deep reinforcement learning in large discrete action spaces. arXiv preprint arXiv:1512.07679.

Ferguson, B., Nagaraj, V., Kara, E.C., Alizadeh, M., 2018. Optimal Planning of Workplace Electric Vehicle Charging Infrastructure with Smart Charging Opportunities. IEEE Conf. Intell. Transp. Syst. Proc. ITSC 2018-Novem, 1149–1154.

Gao, Z., Gao, Y., Hu, Y., Jiang, Z., Su, J., 2020. Application of Deep Q-Network in Portfolio Management. In: 2020 5th IEEE International Conference on Big Data Analytics. ICBDA, IEEE, pp. 268–275. http://dx.doi.org/10.1109/ICBDA49040.2020.9101333, URL https://ieeexplore.ieee.org/document/9101333/.

Ghiassi-Farrokhfal, Y., Rosenberg, C., Keshav, S., Adjaho, M.B., 2016. Joint Optimal Design and Operation of Hybrid Energy Storage Systems. IEEE J. Sel. Areas Commun. 34 (3), 639–650. http://dx.doi.org/10.1109/JSAC.2016.2525599.

Gijsbrechts, J., Boute, R.N., Van Mieghem, J.A., Zhang, D.J., 2022. Can Deep Reinforcement Learning Improve Inventory Management? Performance on Lost Sales, Dual-Sourcing, and Multi-Echelon Problems. Manuf. Serv. Oper. Manag. 24 (3), 1349–1368. http://dx.doi.org/10.1287/msom.2021.1064, URL https://pubsonline.informs.org/doi/10.1287/msom.2021.1064.

Glaessgen, E., Stargel, D., 2012. The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles. In: 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference. (April), American Institute of Aeronautics and Astronautics, Reston, Virigina, pp. 1–14. http://dx.doi.org/10.2514/6.2012-1818.

Grieves, M., Vickers, J., 2017. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In: Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches. Springer International Publishing, pp. 85–113. http://dx.doi.org/10.1007/978-3-319-38756-7_4.

Griffin, A., Hauser, J.R., 1993. The Voice of the Customer. Mark. Sci. 12 (1), 1–27. http://dx.doi.org/10.1287/mksc.12.1.1.

Gust, G., Brandt, T., Mashayekh, S., Heleno, M., DeForest, N., Stadler, M., Neumann, D., 2021. Strategies for microgrid operation under real-world conditions. European J. Oper. Res. 292 (1), 339–352. http://dx.doi.org/10.1016/j.ejor.2020.10.041.

Haarnoja, T., Zhou, A., Abbeel, P., Levine, S., 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: International Conference on Machine Learning. PMLR, pp. 1861–1870.

Hannah, L.A., 2015. Stochastic Optimization. In: International Encyclopedia of the Social & Behavioral Sciences, Second Edi vol. 23, Elsevier, pp. 473–481. http://dx.doi.org/10.1016/B978-0-08-097086-8.42010-6.

He, L., Mak, H.Y., Rong, Y., Shen, Z.J.M., 2017. Service region design for urban electric vehicle sharing systems. Manuf. Serv. Oper. Manag. 19 (2), 309–327. http://dx.doi.org/10.1287/msom.2016.0611.

van Hezewijk, L., Dellaert, N., Van Woensel, T., Gademann, N., 2022. Using the proximal policy optimisation algorithm for solving the stochastic capacitated lot sizing problem. Int. J. Prod. Res. http://dx.doi.org/10.1080/00207543.2022.2056540.

Hoover, Z., Polymeneas, E., Sahdev, S., 2021. How Charging in Buildings Can Power Up the Electric-Vehicle Industry. Technical Report, McKinsey & Company, pp. 1–8, URL https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/how-charging-in-buildings-can-power-up-the-electric-vehicle-industry.

Huang, Y., Zhou, Y., 2015. An optimization framework for workplace charging strategies. Transp. Res. Part C: Emerg. Technol. 52, 144–155. http://dx.doi.org/10.1016/j.trc.2015.01.022.

Jones, D., Snider, C., Nassehi, A., Yon, J., Hicks, B., 2020. Characterising the Digital Twin: A systematic literature review. CIRP J. Manuf. Sci. Technol. 29, 36–52. http://dx.doi.org/10.1016/j.cirpj.2020.02.002.

Jun, M., Meintz, A., 2018. Workplace Charge Management with Aggregated Building Loads. In: 2018 IEEE Transportation and Electrification Conference and Expo, ITEC 2018. IEEE, pp. 519–524. http://dx.doi.org/10.1109/ITEC.2018.8450227.

Kazemi, M.A., Sedighizadeh, M., Mirzaei, M.J., Homaee, O., 2016. Optimal siting and sizing of distribution system operator owned EV parking lots. Appl. Energy 179, 1176–1184. http://dx.doi.org/10.1016/j.apenergy.2016.06.125.

Kazhamiaka, F., Rosenberg, C., Keshav, S., 2019. Tractable lithium-ion storage models for optimizing energy systems. Energy Informatics 2 (1), http://dx.doi.org/10.1186/s42162-019-0070-6.

Ketter, W., Schroer, K., Valogianni, K., 2023. Information Systems Research for Smart Sustainable Mobility: A Framework and Call for Action. Inf. Syst. Res. 34 (3), 1045–1065. http://dx.doi.org/10.1287/isre.2022.1167, http://pubsonline.informs.org/doi/10.1287/isre.2022.1167. https://pubsonline.informs.org/doi/10.1287/isre.2022.1167.

Konda, V., Tsitsiklis, J., 1999. Actor-critic algorithms. Adv. Neural Inf. Process. Syst. 12.

Lee, J.H., Chakraborty, D., Hardman, S.J., Tal, G., 2020. Exploring electric vehicle charging patterns: Mixed usage of charging infrastructure. Transp. Res. Part D: Transp. Environ. 79, 102249. http://dx.doi.org/10.1016/j.trd.2020.102249, URL https://linkinghub.elsevier.com/retrieve/pii/S136192091831099X.

Lee, Z.J., Chang, D., Jin, C., Lee, G.S., Lee, R., Lee, T., Low, S.H., 2018. Large-scale adaptive electric vehicle charging. 2018 IEEE Glob. Conf. Signal Inf. Process. Glob. 2018 - Proc. 863–864. http://dx.doi.org/10.1109/GlobalSIP.2018.8646472.

Lee, Z.J., Li, T., Low, S.H., 2019. ACN-Data: Analysis and applications of an open EV charging dataset. E- Energy 2019 - Proc. the 10th ACM Int. Conf. Futur. Energy Syst. 139–149. http://dx.doi.org/10.1145/3307772.3328313.

Li, S., Xie, F., Huang, Y., Lin, Z., Liu, C., 2020. Optimizing workplace charging facility deployment and smart charging strategies. Transp. Res. Part D: Transp. Environ. 87, 102481. http://dx.doi.org/10.1016/j.trd.2020.102481.

Lillicrap, T., 2015. Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M., 2013. Playing Atari with Deep Reinforcement Learning. pp. 1–9, URL http://arxiv.org/abs/1312.5602. arXiv:1312.5602.

Mukherjee, J.C., Gupta, A., 2015. A Review of Charge Scheduling of Electric Vehicles in Smart Grid. IEEE Syst. J. 9 (4), 1541–1553. http://dx.doi.org/10.1109/JSYST.2014.2356559.

Neon Neue Energieökonomik and Technical University of Berlin and ETH Zürich and DIW Berlin, 2024. Open power system data. URL https://open-power-system-data.org/.

Nunes, P., Figueiredo, R., Brito, M.C., 2016. The use of parking lots to solar-charge electric vehicles. Renew. Sustain. Energy Rev. 66, 679–693. http://dx.doi.org/10.1016/j.rser.2016.08.015.

Panzer, M., Bender, B., 2022. Deep reinforcement learning in production systems: a systematic literature review. Int. J. Prod. Res. 60 (13), 4316–4341. http://dx.doi.org/10.1080/00207543.2021.1973138.

Powell, W.B., 2014. Energy and uncertainty: Models and algorithms for complex energy systems. AI Mag. 35 (3), 8–21. http://dx.doi.org/10.1609/aimag.v35i3.2540.

Qi, W., Shen, Z.J.M., 2018. A Smart-City Scope of Operations Management. Prod. Oper. Manage. 1–14. http://dx.doi.org/10.1111/poms.12928.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20 (C), 53–65. http://dx.doi.org/10.1016/0377-0427(87)90125-7.

Schleich, B., Anwer, N., Mathieu, L., Wartzack, S., 2017. Shaping the digital twin for design and production engineering. CIRP Ann 66 (1), 141–144. http://dx.doi.org/10.1016/j.cirp.2017.04.040.

Sharifi, P., Banerjee, A., Feizollahi, M.J., 2020. Leveraging owners' flexibility in smart charge/discharge scheduling of electric vehicles to support renewable energy integration. Comput. Ind. Eng. 149 (July), 106762. http://dx.doi.org/10.1016/j.cie.2020.106762.

Sierla, S., Kyrki, V., Aarnio, P., Vyatkin, V., 2018. Automatic assembly planning based on digital product descriptions. Comput. Ind. 97, 34–46. http://dx.doi.org/10.1016/j.compind.2018.01.013.

Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., Hassabis, D., 2016. Mastering the game of Go with deep neural networks and tree search. Nature 529 (7587), 484–489. http://dx.doi.org/10.1038/nature16961, URL http://dx.doi.org/10.1038/nature16961.

Sutton, R.S., 2019. The Bitter Lesson. URL http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: An Introduction, second ed. The MIT Press.

Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., Sui, F., 2018. Digital twin-driven product design, manufacturing and service with big data. Int. J. Adv. Manuf. Technol. 94 (9–12), 3563–3576. http://dx.doi.org/10.1007/s00170-017-0233-1.

van der Valk, H., Haße, H., Möller, F., Arbter, M., Henning, J.L., Otto, B., 2020. A Taxonomy of Digital Twins. 26th Am. Conf. Inf. Syst. AMCIS 2020 1–10.

Van Hasselt, H., Guez, A., Silver, D., 2016. Deep reinforcement learning with double q-learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 30, (1).

Wang, H., Odoni, A., 2016. Approximating the Performance of a "Last Mile" Transportation System. Transp. Sci. 50 (2), 659–675. http://dx.doi.org/10.1287/trsc.2014.0553.

Wu, D., Zeng, H., Lu, C., Boulet, B., 2017. Two-Stage Energy Management for Office Buildings with Workplace EV Charging and Renewable Energy. IEEE Trans. Transp. Electrification 3 (1), 225–237. http://dx.doi.org/10.1109/TTE.2017.2659626.

Xie, J., Liu, Y., Chen, N., 2023. Two-Sided Deep Reinforcement Learning for Dynamic Mobility-on-Demand Management with Mixed Autonomy. Transp. Sci. 57 (4), 1019–1046. http://dx.doi.org/10.1287/trsc.2022.1188, URL https://pubsonline.informs.org/doi/10.1287/trsc.2022.1188.