

American Journal of Data Science and Analytics

Volume: 7; **Issue:** 1
Pages: 405-441

6G-Enabled Multi-Agent Deep Reinforcement Learning for Intelligent Vehicle-to-Grid Energy Optimization and Renewable Integration in Smart Grids: A Data-Driven Analytics Framework and Simulation Study

Abdur Rashid[1];

[1]. *Junior Project Coordinator, Virtual American Companies, LLC, Texas, USA;*
M.S. Engineering Management, Westcliff University, USA
Email: rabdurrashid091@gmail.com

Doi: 10.63125/7w76dg74
Received: *11 October 2025;* ***Revised:*** *20 November 2025;* ***Accepted:*** *25 December 2025;* ***Published:*** *18 January 2026;*

Abstract

Vehicle-to-grid (V2G) coordination asks hundreds of charging hubs to decide, every few minutes, how much to charge or discharge the vehicles connected to them so that energy cost, peak demand, transformer loading, battery wear and users' departure needs are balanced while renewable generation is absorbed. Multi-agent deep reinforcement learning (MARL) is a natural control architecture for this problem, and sixth-generation (6G) wireless systems promise the ultra-reliable low-latency links and edge intelligence it needs, but the value of that connectivity has rarely been quantified. This article designs a data-driven analytics framework that links a physical feeder-and-fleet model, a communication layer, a multi-agent learning layer and a decision layer, and evaluates it on a transparent simulation testbed: a radial feeder with eight charging hubs and 200 vehicles behind a 1,200 kW transformer, hub-level photovoltaics and feeder-level wind, duck-curve prices and carbon intensity, and three communication archetypes (4G-LTE, 5G, 6G) that differ in message loss, setpoint delay, fleet-flexibility messaging and eligibility for fast reserve products. Five policies are compared on 40 held-out days: uncontrolled charging, a time-of-use heuristic, independent deep Q-learning, value-decomposition MARL and a perfect-information linear-programming bound. Every coordinated policy removed the transformer overload of uncontrolled charging, cutting daily cost by 27–40%; the learned policies beat a capacity-aware heuristic only when the network supplied fleet-flexibility messages and reserve eligibility (by 2.5–7.8% under 6G, not under 4G) while remaining 30–41% above the bound. The value of 6G over 4G to the learned policy was 613 per vehicle-year gross and 552 net of a connectivity premium, positive in every economic draw, split between a market channel (reserve eligibility) and an information channel (fleet messages), with link loss and delay worth little and the reserve price explaining 87% of its variance. Savings rose with photovoltaics and price volatility, and the policy did not generalise to a feeder without photovoltaics. All inputs are stylised with stated ranges; the testbed and results file are provided.

Keywords

Vehicle-To-Grid, Multi-Agent Reinforcement Learning, 6G, Ultra-Reliable Low-Latency Communication, Age Of Information, Renewable Integration, Smart Grid, Demand Response, Decision Analytics, Simulation

Introduction

Electric vehicles are becoming the largest flexible load on distribution networks, and their batteries, parked for most of the day, are the largest distributed storage resource that a grid operator could call on (Kempton & Tomić, 2005a; Kempton & Tomić, 2005b; IEA, 2024). Vehicle-to-grid (V2G) operation (bidirectional charging that lets a fleet absorb surplus renewable generation, shave peaks and provide reserves) has been studied for two decades, and its economics, its effects on distribution networks and its social dimensions are reviewed in a large literature (Lopes et al., 2011; Yilmaz & Krein, 2013; Tan et al., 2016; Mwasilu et al., 2014; Sovacool et al., 2018). What has changed is scale and timing: the coordination problem now involves hundreds of charging hubs and thousands of vehicles whose arrivals, departures and energy needs are uncertain, whose charging decisions interact through transformer limits, voltages and tariffs, and whose value to the grid depends on reacting within seconds to prices, renewable output and frequency events (Hu et al., 2016; Mukherjee & Gupta, 2015; Han et al., 2010). Centralised optimisation of this problem is exact but brittle; decentralised heuristics are robust but leave value on the table; and multi-agent reinforcement learning (MARL), in which each hub learns a policy from its own observations while a central critic coordinates training, has emerged as the architecture that promises both (Vázquez-Canteli & Nagy, 2019; da Silva et al., 2020; Shin et al., 2020; Zhang et al., 2021).

Two gaps separate that promise from practice, and both are data problems. The first is communication. A MARL controller is only as good as the information its agents receive: the aggregator's view of feeder load, the fleet-flexibility messages that let one hub know what the others can do, and the setpoints that the hubs execute. Fourth-generation cellular links deliver these with delays and losses that make the information stale; fifth-generation ultra-reliable low-latency communication (URLLC) and the sixth-generation (6G) vision of sub-millisecond latency, 10^{-5} reliability and intelligence at the edge promise to remove the staleness and to make fast reserve products available to vehicle fleets (Saad et al., 2020; Tataria et al., 2021; Letaief et al., 2019; Bennis et al., 2018; Popovski et al., 2019; Noor-A-Rahim et al., 2022). Yet the 6G literature describes capabilities, not decisions: it does not say how much a hub operator's cost or a feeder's peak changes when message age falls from minutes to milliseconds, nor whether that change pays for the connectivity. The second gap is evaluation. Deep reinforcement learning results are sensitive to seeds, hyperparameters and evaluation protocol (Henderson et al., 2018; Agarwal et al., 2021; Dulac-Arnold et al., 2021), and many V2G learning studies report a single training run against a weak baseline without a bound on what was achievable (Sumaiya, 2024, 2025, 2026). A decision-maker (a fleet operator, a distribution utility, an aggregator) needs a probability that the controller beats the alternatives, the inputs that drive the outcome, and the value of the connectivity it would have to buy.

This article designs a data-driven decision-analytics framework for 6G-enabled MARL V2G coordination and evaluates it on a transparent simulation testbed. The framework, in the tradition of design-science research (Hevner et al., 2004; Peffers et al., 2007; Gregor & Hevner, 2013), has four layers: a physical layer (feeder, transformer, hubs, fleets, renewables, prices), a communication and edge layer (telemetry, setpoints, fleet-flexibility messages, communication archetypes, reserve eligibility), a multi-agent learning layer (observations, a safety layer, independent and value-decomposition Q-learning, baselines and a bound) and a decision layer (metrics, Monte Carlo economics, scenarios, governance). The testbed is a radial feeder with eight charging hubs and 200 vehicles behind a 1,200 kW transformer, hub-level photovoltaics and feeder-level wind, duck-curve prices and a carbon-intensity profile that follows solar output; three communication archetypes represent 4G-LTE, 5G and 6G links by message loss, setpoint delay, whether fleet-flexibility messages are available and what share of fast reserve products the link qualifies for; five policies (uncontrolled charging, a time-of-use heuristic, independent deep Q-learning (IQL), value-decomposition MARL (VDN) and a perfect-information linear-programming oracle) are trained with three seeds and evaluated on 40 held-out days. The testbed's purpose is not to predict any real fleet's economics (it uses no measured data) but to show how the framework's outputs behave: which policy wins with what probability, how much the communication layer is worth and through which channels, how the answer changes with EV penetration, renewable share and price volatility, and which economic inputs drive it. Because every input is a parameter, an operator can replace the ranges with its own data; the testbed and the complete results file accompany

the article.

Four research questions structure the study. RQ1: what do the V2G, charging-coordination, MARL, 6G and smart-grid-communication literatures require of a decision framework for connected V2G control, and what framework meets them? RQ2: how do uncontrolled, heuristic, independent and cooperative learning policies compare on cost, peak demand, grid constraints, user service, renewable integration and emissions, and how close do they come to the perfect-information bound? RQ3: what is the value of the communication layer (of low loss, low delay, fleet-flexibility messages and reserve eligibility) and how does it decompose, degrade and depend on economic inputs? RQ4: how do the results generalise across EV penetration, renewable share and price volatility, and what do they imply for implementation, standards and governance?

The results support a conditional answer. On the testbed, coordination of any kind removes the transformer overload that uncontrolled charging causes and cuts daily cost by 27–40%; learned policies beat a capacity-aware time-of-use heuristic only when the network gives them fleet-flexibility messages and reserve eligibility (under 6G, not under 4G) and the independent learner with individual rewards did better than value decomposition at the training budget used, while both remained 30–41% above the perfect-information bound. The value of 6G over 4G to the learned policy, 613 per vehicle-year gross and 552 net of a connectivity premium on the testbed, comes in roughly equal parts from market access (fast-reserve eligibility) and from the edge aggregator's fleet-flexibility messages, and hardly at all from the loss and delay of the radio link itself; it is positive in every economic draw, but 87% of its variance comes from one input, the reserve price. Savings rise with photovoltaic scale and price volatility, the learned policy does not generalise to a feeder without photovoltaics, and at high penetration the transformer binds for every policy but the bound. The limitation to state at the outset is that every input is stylised (the communication archetypes in particular are representations of what the standards literature describes, not measurements of any network) and the article reports the sensitivity of each conclusion to the ranges assumed.

Framing the problem as decision analytics rather than as control alone changes what the analysis has to produce. The analytics literature distinguishes prediction from decision and shows that connecting models to the decisions they serve matters more than marginal accuracy (Bertsimas & Kallus, 2020; Elmachtoub & Grigas, 2022); reinforcement learning is decision-focused by construction, but its outputs are policies and returns, not probabilities, break-evens and values of information, which is what a capital decision about connectivity needs (Howard, 1966; Morgan & Henrion, 1990). The behavioural literature warns that single-run results anchor judgement (Tversky & Kahneman, 1974; Kahneman, 2011), and the reinforcement-learning community has shown how much seeds and protocols move reported numbers (Henderson et al., 2018; Agarwal et al., 2021). A framework for connected V2G therefore has to state its data and their provenance, evaluate with seeds and held-out days against a bound, propagate economic uncertainty into probabilities, present results as ranges (Spiegelhalter, 2017), and be interpretable enough that an operator can see why a policy acts as it does (Rudin, 2019).

The remainder of the article reviews the literature and derives the requirements (RQ1), describes the testbed, the learning agents and the evaluation methods, reports outcomes, the value of connectivity, generalisation and economics (RQ2–RQ4), and discusses implications, analytics practice, limitations and further research. Twelve figures and fifteen tables carry the detail.

Background and Related Work

Vehicle-to-grid: fundamentals, economics and grid impacts

The V2G concept was set out by Kempton and Tomić, who calculated the capacity and net revenue that vehicle fleets could offer to power markets and argued that regulation and spinning reserves, not bulk energy, are the markets in which battery degradation and infrastructure cost are repaid (Kempton & Tomić, 2005a; Kempton & Tomić, 2005b); the conceptual framework for an aggregator that mediates between vehicles and markets followed (Guille & Gross, 2009), and aggregator designs for frequency regulation and market bidding were developed (Han et al., 2010; Vagropoulos & Bakirtzis, 2013). The grid-side literature quantified the other half of the problem: uncoordinated charging raises residential feeder peaks and losses and can overload transformers, while coordinated charging flattens them (Clement-Nyns et al., 2010; Lopes et al., 2011; Hilshey et al., 2013; IEEE, 2011); reviews of modelling approaches and of V2G effects on distribution systems and utility interfaces consolidate these findings

(Richardson, 2013; Yilmaz & Krein, 2013). The economics of using vehicle batteries for grid services depend on degradation, and the evidence is mixed: early analyses found bulk arbitrage uneconomic because of battery wear (Peterson et al., 2010), bidirectional cycling has been shown to accelerate ageing under some protocols (Dubarry et al., 2017), while optimised V2G that keeps vehicles in favourable state-of-charge windows can extend battery life (Uddin et al., 2017); degradation models for cell-life assessment make the trade-off computable (Xu et al., 2018). Reviews of V2G technologies and optimisation techniques (Tan et al., 2016; Mwasilu et al., 2014), of fleet management services (Hu et al., 2016) and of the neglected social dimensions of a V2G transition (Sovacool et al., 2018; Noel et al., 2019) complete the picture, and the integration interface is now standardised in ISO 15118-20 for bidirectional sessions and IEEE 1547-2018 for interconnection (ISO, 2022; IEEE, 2018).

Charging coordination: optimisation, decentralised protocols and heuristics

Charging coordination has an optimisation literature. Optimal charging strategies for unidirectional and bidirectional V2G were formulated as linear and quadratic programmes with energy and ancillary-service markets (Sortomme & El-Sharkawi, 2011; Sortomme & El-Sharkawi, 2012), and reviews of charge scheduling catalogue the objectives (cost, peak, losses, voltage, battery life) and the solution methods (Mukherjee & Gupta, 2015). Because centralised solution of large fleets is impractical and intrusive, decentralised protocols were developed in which vehicles respond to broadcast prices and converge to the centralised optimum (Ma et al., 2013; Gan et al., 2013), in the tradition of controllable-load aggregation and state estimation of load populations (Callaway & Hiskens, 2011; Mathieu et al., 2013). Demand-response reviews place V2G among the flexible loads that tariffs and signals can shape (Palensky & Dietrich, 2011; Siano, 2014; Deng et al., 2015; O'Connell et al., 2014), with real-time pricing as the efficient but behaviourally demanding instrument (Borenstein, 2005). Open datasets and simulators of workplace charging (ACN-Data and ACN-Sim) supply the arrival, departure and energy distributions on which such studies now rest (Lee et al., 2019; Lee et al., 2021), and CityLearn provides a gym environment for demand-response learning (Vázquez-Canteli et al., 2019).

Deep and multi-agent reinforcement learning

Deep Q-networks showed that a neural network could learn control from raw observations with experience replay and target networks (Mnih et al., 2015); policy-gradient and actor-critic methods for continuous control followed (Lillicrap et al., 2015; Schulman et al., 2017); the textbook treatment is Sutton and Barto (Sutton & Barto, 2018). Multi-agent learning adds non-stationarity (each agent's environment includes the other learners) and the independent-learner baseline dates from Tan (Tan, 1993); surveys of multi-agent reinforcement learning before and after the deep era map the design space of cooperation, communication and credit assignment (Buşoniu et al., 2008; Hernández-Leal et al., 2019; Zhang et al., 2021; Oroojlooy & Hajinezhad, 2023). The centralised-training, decentralised-execution paradigm resolves the non-stationarity by training with global information while executing locally: multi-agent actor-critic (Lowe et al., 2017), counterfactual policy gradients (Foerster et al., 2018), value-decomposition networks that sum per-agent action values into a team value (Sunehag et al., 2018), their monotonic generalisation QMIX (Rashid et al., 2018), and multi-agent PPO (Yu et al., 2022). The evaluation literature is a warning: results depend on seeds, hyperparameters and reporting, and few runs with point estimates mislead (Henderson et al., 2018; Agarwal et al., 2021); the challenges of real-world reinforcement learning (safety constraints, partial observability, delays, non-stationarity) are catalogued by Dulac-Arnold et al. (Dulac-Arnold et al., 2021).

Reinforcement learning for EV charging and power systems

Applications of reinforcement learning to demand response, EV charging and power systems are reviewed from several angles (Vázquez-Canteli & Nagy, 2019; Yang et al., 2020; Perera & Kamalaruban, 2021; Cao et al., 2020; Chen et al., 2022; Zhang et al., 2020; Rolf et al., 2023). For EV charging specifically, model-free deep reinforcement learning schedules single-station charging against uncertain prices (Wan et al., 2019), safe reinforcement learning enforces charging constraints (Li et al., 2020), fleet charging in day-ahead markets was learned with heuristics (Vandael et al., 2015), model-free coordination was defined and evaluated against optimisation (Sadeghianpourhamami et al., 2020), multi-agent learning coordinates charging across stations (da Silva et al., 2020), and cooperative multi-agent deep reinforcement learning manages PV and storage at charging stations (Shin et al., 2020). The common findings are that learned policies approach optimisation benchmarks when observations are

informative (Aminul Islam, 2024; Aminul Islam & Md, 2023; Md. Morshedul & Aminul Islam, 2022), that constraints need a safety layer, and that coordination across agents requires either shared rewards or communicated state, which is where the communication layer enters (Kynatun & Saba, 2024; Mansura Akter & Sumaiya, 2022).

6G, URLLC, age of information and V2X

The 6G vision papers describe a network of sub-millisecond latency, ultra-high reliability, native artificial intelligence, integrated sensing and communication, and intelligence at the edge (Saad et al., 2020; Tataria et al., 2021; Letaief et al., 2019; Dang et al., 2020; Zhang et al., 2019; Giordani et al., 2020); the ITU-R framework for IMT-2030 sets out the usage scenarios and capabilities that standardisation will target (ITU, 2023). URLLC is the capability that control applications need, and its tail, risk and scale properties are the subject of a foundational analysis (Bennis et al., 2018) and of a design review (Popovski et al., 2019). Age of information (the time since the last received update was generated) is the metric that connects link quality to control quality (Kaul et al., 2012; Yates et al., 2021), and distributed learning over wireless networks, including federated and multi-agent learning at the edge, has its own research agenda (Chen et al., 2021; Park et al., 2019). Vehicle-to-everything (V2X) communication is the concrete setting: 5G NR V2X is tutorially described (Castañeda Garcia et al., 2021) and the 6G V2X roadmap sets out the enabling technologies and open problems (Noor-A-Rahim et al., 2022). Smart-grid communications have their own survey literature on requirements, standards and infrastructures (Gungor et al., 2011; Yan et al., 2013; Fang et al., 2012; Kalalas et al., 2016), and the role of 5G in smart grids and demand response has been reviewed (Reka et al., 2019; Hui et al., 2020). What this literature does not provide is a decision model: the link between message loss, delay and reserve eligibility on one side and fleet cost, peaks and emissions on the other.

Renewable integration, flexibility and distribution networks

High shares of variable renewables require flexibility (storage, demand response, transmission) and vehicle fleets are among the largest potential sources (Denholm & Hand, 2011; Lund et al., 2015; Jenkins et al., 2018); the market value of variable renewables falls as their share rises, which is the price signal that flexible charging responds to (Hirth, 2013). Validated reanalysis-based PV and wind profiles are the standard inputs for such studies (Pfenninger & Staffell, 2016; Staffell & Pfenninger, 2016), and the carbon intensity of the grid, published by region and projected to fall, determines the emissions value of shifting load (EPA, 2024; Gagnon et al., 2021; Siler-Evans et al., 2012). Distribution-network effects are computed on test feeders (Baran & Wu, 1989; Kersting, 1991) with power-flow tools (Thurner et al., 2018; Zimmerman et al., 2011) or linearised branch-flow models (Farivar & Low, 2013); electrification of transport and heat shifts peaks and raises demand charges (Waite & Modi, 2020; Buonocore et al., 2022; Mai et al., 2018).

Decision analytics and interpretable learning

The management-science literature supplies the framework's remaining requirements: prescriptive analytics connects models to decisions (Bertsimas & Kallus, 2020; Elmachtoub & Grigas, 2022; Ban & Rudin, 2019); aleatoric uncertainty (future prices, weather, arrivals) is distinguished from epistemic uncertainty (costs, degradation, link quality) because only the second can be bought down (Hüllermeier & Waegeman, 2021; Kendall & Gal, 2017); risk is identified, analysed, treated and recorded (ISO, 2018) and communicated as probabilities and ranges (Spiegelhalter, 2017); high-stakes recommendations rest on inspectable models (Rudin, 2019; Lipton, 2018; Doshi-Velez & Kim, 2017), with importance, partial dependence and shallow trees making flexible models readable (Lundberg & Lee, 2017; Lundberg et al., 2020; Fisher et al., 2019; Molnar, 2025); and human judgement is supported rather than replaced (Amershi et al., 2019; Shneiderman, 2020; Fildes et al., 2009). Variance-based sensitivity analysis and value of information connect uncertainty to decisions (Sobol', 2001; Saltelli et al., 2010; Herman & Usher, 2017; Howard, 1966), simulation studies follow conventions of verification and common random numbers (Law, 2015; Sargent, 2013), and analytics capability pays through organisational routines and data quality (Mikalef et al., 2019; Brynjolfsson et al., 2021; Hazen et al., 2014; Haug et al., 2011).

Requirements and the framework

Table 1 organises the literature into nine strands and the gap each leaves. Six requirements follow. R1, a physical model with constraints: transformer rating, feeder voltages, hub capacity, battery limits and users' departure targets must bind, because the value of coordination is the value of respecting them

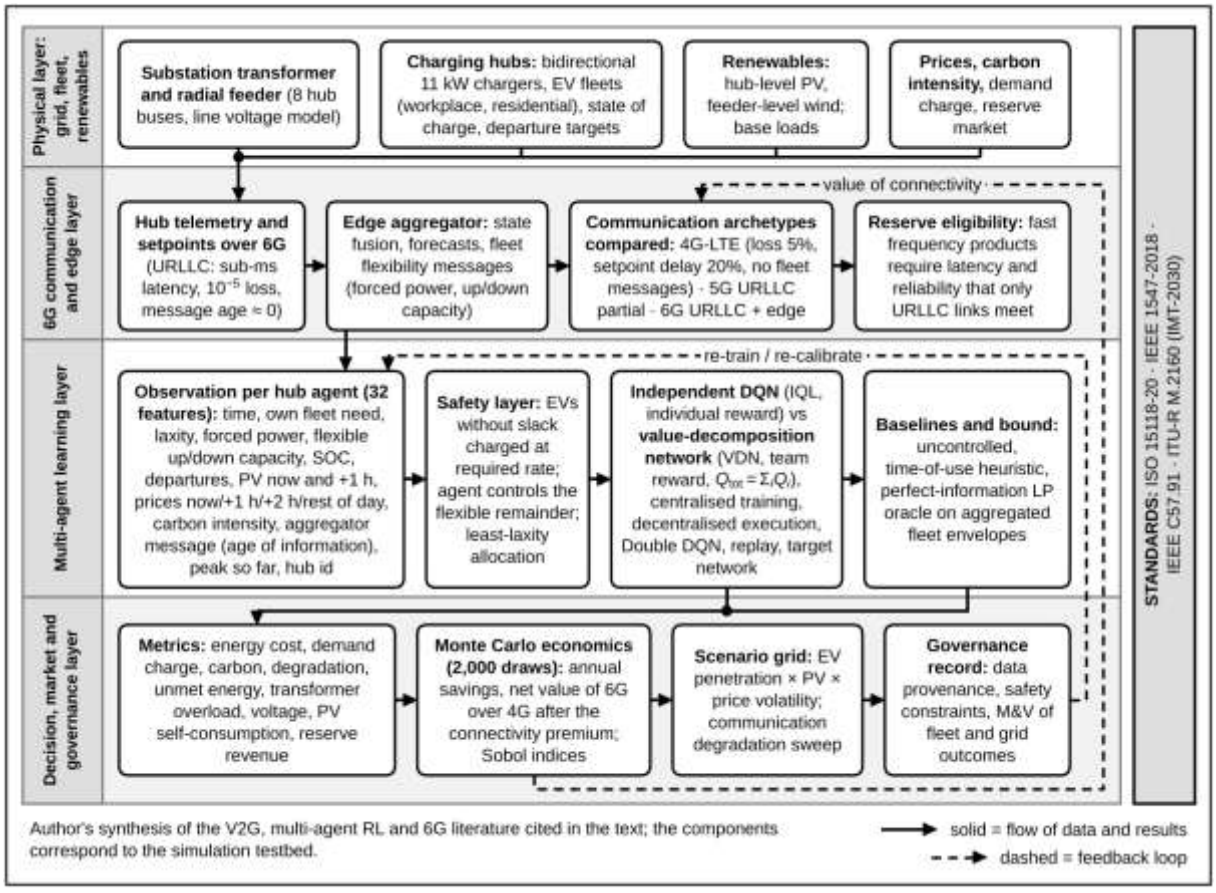
cheaply. R2, communication as a modelled layer: message loss, delay, information age, what is communicated and which market products the link qualifies for must be parameters, so that connectivity has a value. R3, learning with safety and coordination: agents must not violate user service or hard limits while learning, and cooperation must be achievable without centralised execution.

Table 1: Literature strands, what they establish, representative sources and the gap addressed

Strand	What the literature establishes	Representative sources	Gap addressed here
V2G fundamentals and economics	Fleets can supply capacity and reserves; regulation and reserve markets, not bulk energy, repay degradation and infrastructure; aggregators mediate; degradation evidence is mixed and depends on the state-of-charge window	Kempton & Tomić (2005a, 2005b); Guille & Gross (2009); Han et al. (2010); Vagropoulos & Bakirtzis (2013); Peterson et al. (2010); Dubarry et al. (2017); Uddin et al. (2017); Xu et al. (2018)	Reserve revenue, degradation and user service inside one reward and one economic model
Grid impacts of EV charging	Uncoordinated charging raises feeder peaks, losses and transformer loading; coordinated charging flattens them; interconnection and transformer-loading standards bound the problem	Clement-Nyns et al. (2010); Lopes et al. (2011); Hilshey et al. (2013); IEEE (2011, 2018); Richardson (2013); Yilmaz & Krein (2013)	Transformer, hub and voltage constraints as penalties the agents must learn to respect
Charging coordination: optimisation and decentralised protocols	Linear and quadratic programmes for V2G scheduling; decentralised price-based protocols converge to the centralised optimum; load-population aggregation and state estimation	Sortomme & El-Sharkawi (2011, 2012); Mukherjee & Gupta (2015); Ma et al. (2013); Gan et al. (2013); Callaway & Hiskens (2011); Mathieu et al. (2013); Vandael et al. (2015)	Perfect-information LP as a bound on what any coordination can achieve
Demand response, tariffs and open charging data	Flexible loads respond to tariffs and signals; real-time pricing is efficient but demanding; open workplace-charging data and simulators supply session distributions	Palensky & Dietrich (2011); Siano (2014); Deng et al. (2015); O'Connell et al. (2014); Borenstein (2005); Lee et al. (2019, 2021); Vázquez-Canteli et al. (2019)	Session ranges, duck-curve prices and demand charges as testbed parameters with stated ranges
Deep and multi-agent reinforcement learning	DQN with replay and target networks; actor-critic methods; non-stationarity of independent learners; centralised training with decentralised execution (MADDPG, COMA, VDN, QMIX, MAPPO); seed and protocol sensitivity	Mnih et al. (2015); Lillicrap et al. (2015); Schulman et al. (2017); Sutton & Barto (2018); Tan (1993); Buşoniu et al. (2008); Lowe et al. (2017); Foerster et al. (2018); Sunehag et al. (2018); Rashid et al. (2018); Yu et al. (2022); Henderson et al. (2018); Agarwal et al. (2021)	IQL and VDN with three seeds, held-out days, interval estimates and a bound
Reinforcement learning for EV charging and power systems	Model-free DRL schedules charging against uncertain prices; safe RL enforces constraints; multi-agent learning coordinates stations and manages PV and storage; reviews map the field	Vázquez-Canteli & Nagy (2019); Yang et al. (2020); Perera & Kamalaruban (2021); Cao et al. (2020); Chen et al. (2022); Zhang et al. (2020); Wan et al. (2019); Li et al. (2020); Sadeghianpourhamami et al. (2020); da Silva et al. (2020); Shin et al. (2020)	Safety layer, communicated state and team reward as explicit design choices whose value is measured
6G, URLLC, age of information and V2X	Sub-millisecond latency, 10^{-5} reliability and edge intelligence as the 6G vision; URLLC tails and risk; age of information links link quality to control quality; 5G/6G V2X roadmaps; smart-grid communication requirements	Saad et al. (2020); Tataria et al. (2021); Letaief et al. (2019); Dang et al. (2020); Zhang et al. (2019); Giordani et al. (2020); ITU (2023); Bennis et al. (2018); Popovski et al. (2019); Kaul et al. (2012); Yates et al. (2021); Noor-A-Rahim et al. (2022); Castañeda Garcia et al. (2021); Gungor et al. (2011); Reka et al. (2019); Hui et al. (2020)	Communication as a modelled layer (loss, delay, messages, reserve eligibility) with a measured value
Renewable integration, flexibility and distribution networks	High renewable shares need flexibility; the market value of variable renewables falls with share; validated PV and wind profiles; carbon intensity by region; linearised branch-flow and test feeders; electrification raises peaks and demand charges	Denholm & Hand (2011); Lund et al. (2015); Jenkins et al. (2018); Hirth (2013); Pfenninger & Staffell (2016); Staffell & Pfenninger (2016); EPA (2024); Farivar & Low (2013); Baran & Wu (1989); Waite & Modi (2020)	PV self-consumption, emissions and peak outcomes across a scenario grid of penetration, PV and volatility.
Decision analytics and interpretable learning	Prescriptive analytics connects models to decisions; aleatoric vs epistemic uncertainty; risk recorded and communicated as ranges; interpretable models; Sobol sensitivity; value of information; simulation conventions; analytics capability	Bertsimas & Kallus (2020); Elmachtoub & Grigas (2022); Hüllermeier & Waegeman (2021); Spiegelhalter (2017); Rudin (2019); Sobol' (2001); Saltelli et al. (2010); Howard (1966); Law (2015); Sargent (2013); Mikalef et al. (2019)	Probabilities of improvement, value of connectivity, Sobol indices and a readable metamodel as the framework's outputs

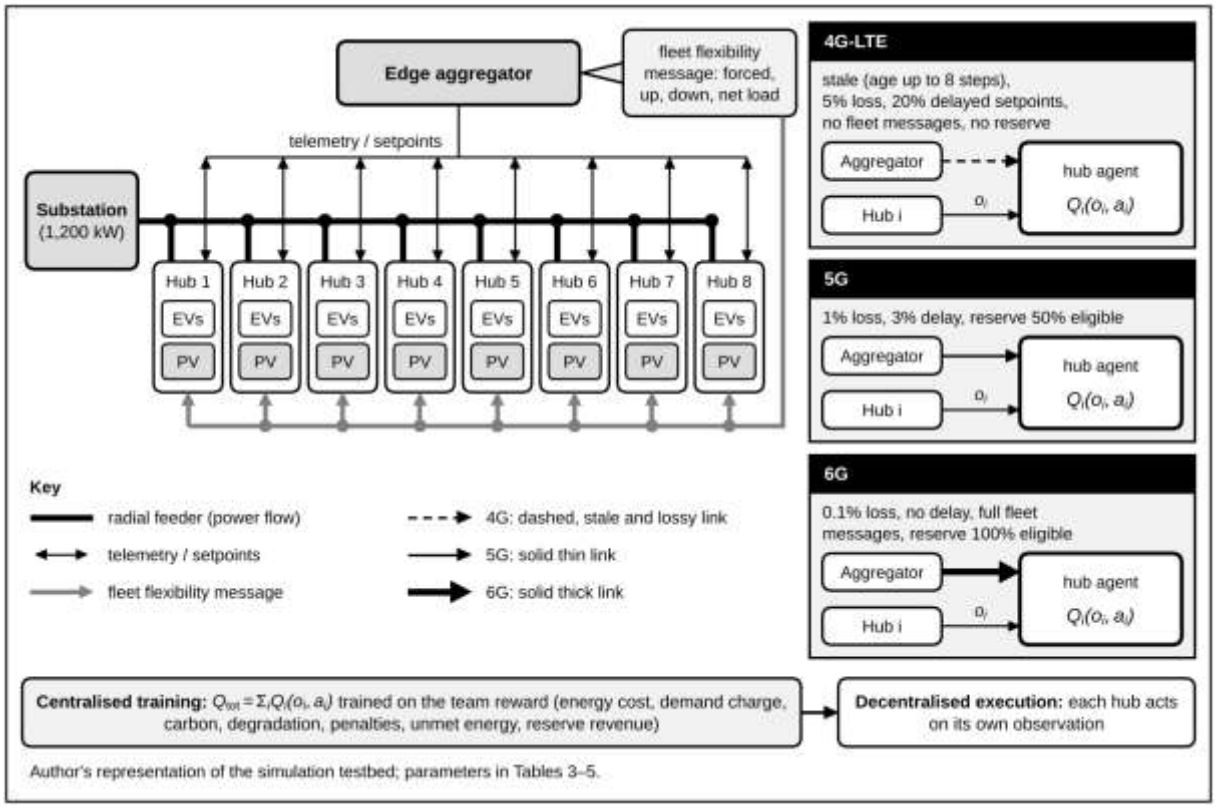
R4, evaluation against a bound with seeds and held-out days: results must be reported as distributions over seeds and days and compared with a perfect-information bound and with simple heuristics. R5, decision outputs: probabilities of improvement, cost decompositions, break-even connectivity premiums, Sobol indices and generalisation across conditions, so that the decision can be to deploy, to upgrade connectivity, or to wait. R6, feedback and governance: measured fleet and grid outcomes must return to the models, and the data provenance, safety constraints and decisions must be recorded. From these, the framework (Figure 1) is designed as four layers: the physical layer, the 6G communication and edge layer, the multi-agent learning layer and the decision, market and governance layer, with feedback from governance to learning and from economics to the communication layer. Figure 2 shows how information flows to the hub agents under each communication archetype and how centralised training with a team value and decentralised execution fit together; the Methods give the models.

Figure 1: 6G-enabled V2G decision-analytics architecture: physical layer, 6G communication and edge layer, multi-agent learning layer, and decision, market and governance layer, with feedback loops



Three design choices distinguish the framework from a controller study. It treats communication as a layer whose parameters have a value, so that the question "is 6G worth it?" has an answer in the operator's units. It evaluates learning against a perfect-information bound and a heuristic with seeds and held-out days, so that the gap to optimality and the variance of the result are reported, not hidden. And it produces probabilities, decompositions and sensitivity indices rather than a single return, so that the decision to deploy, to upgrade connectivity or to wait can be taken under uncertainty. The evaluation that follows tests whether these outputs are informative on a testbed whose truth is known.

Figure 2: Multi-agent learning schematic: feeder, hubs, edge aggregator and the information flows to hub agents under 4G-LTE, 5G and 6G links, with centralised training and decentralised execution

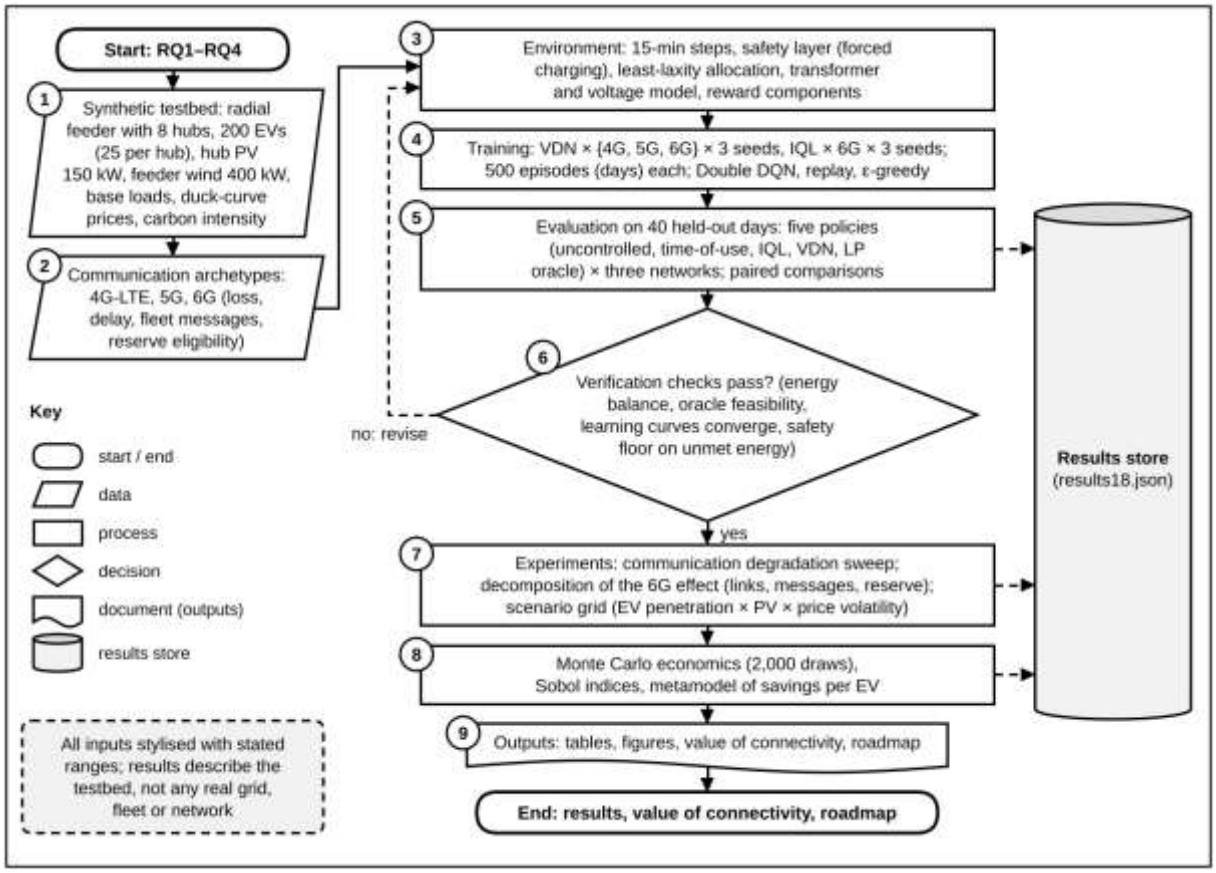


Methods

Study design

The study follows the design-science process of problem and requirements, design of the artefact, evaluation and communication (Peffers et al., 2007; Hevner et al., 2004), with an artificial evaluation on a simulation testbed, chosen because the framework's outputs (probabilities of improvement, the value of connectivity, gaps to a bound) can only be checked where the truth is known (Venable et al., 2016; Sonnenberg & vom Brocke, 2012). Figure 3 shows the design as a flowchart: a synthetic testbed and three communication archetypes, an environment with a safety layer and reward components, training of value-decomposition and independent agents with three seeds each, evaluation of five policies on 40 held-out days, verification, a communication-degradation sweep and a decomposition of the 6G effect, a scenario grid, Monte Carlo economics with Sobol indices and a metamodel, and outputs. The simulation follows the conventions of simulation studies (explicit assumptions, verification of components, common random numbers (the same held-out days for every policy) (Law, 2015; Sargent, 2013)) and the reporting conventions urged for deep reinforcement learning: multiple seeds, interval estimates, and comparison with simple baselines and a bound (Henderson et al., 2018; Agarwal et al., 2021). It is implemented in Python with NumPy and SciPy (the networks are small NumPy multilayer perceptrons trained by hand-written backpropagation), with SALib, LightGBM and scikit-learn for the analysis (Harris et al., 2020; Virtanen et al., 2020; Herman & Usher, 2017; Ke et al., 2017; Pedregosa et al., 2011), runs in about 50 minutes on two cores, and writes a results file from which every table and figure is produced. No measured data from any grid, fleet or network were used.

Figure 3: Study design as a flowchart: testbed and communication archetypes, environment, training, evaluation on held-out days, verification, experiments, Monte Carlo economics and metamodel, outputs

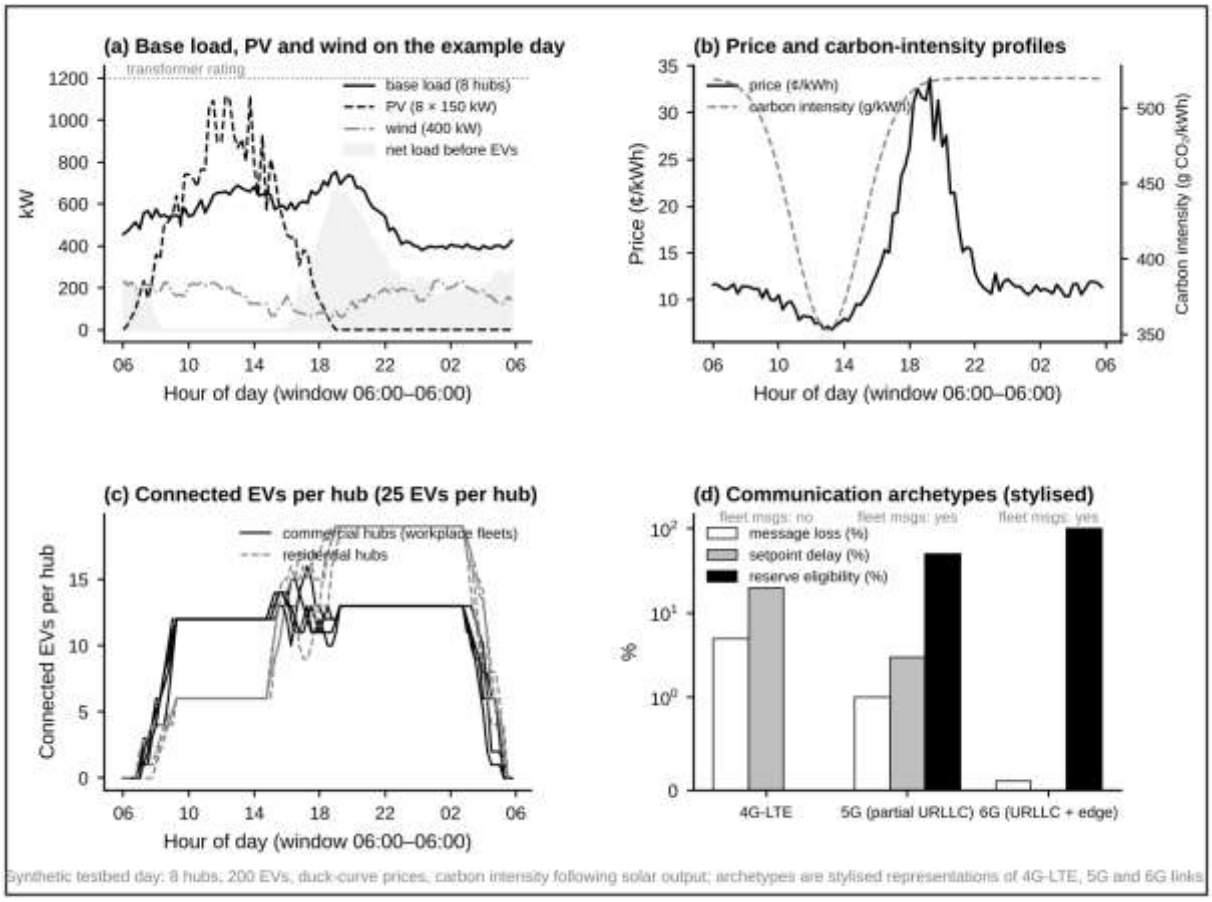


Testbed: feeder, hubs, fleets, renewables and prices

Table 2 lists the testbed and Figure 4 shows an example day. The physical system is a radial feeder with eight charging hubs behind a substation transformer rated at 1,200 kW, with a linearised voltage drop along the feeder computed from the cumulative downstream flow (0.4% of rated flow per hub section, limit 0.95 pu) in the manner of linearised branch-flow models (Farivar & Low, 2013); each hub has a feeder connection limit of 320 kW. Each hub serves 25 vehicles with bidirectional 11 kW chargers at 93% efficiency in each direction, battery capacities of 40–80 kWh, arrival state of charge of 20–60%, a departure target of 85%, and a lower limit of 30% for discharge; commercial hubs (even-numbered) host workplace fleets arriving between 07:00 and 09:30 and leaving between 16:00 and 19:00, residential hubs host fleets arriving between 15:00 and 19:30 and leaving between 03:00 and 05:45 the next morning, in the ranges of the open workplace-charging data (Lee et al., 2019). The simulated day runs from 06:00 to 06:00 in 96 steps of 15 minutes. Each hub has 150 kW of photovoltaics whose daily output follows a clear-sky shape scaled by a cloudiness factor (0.3–1.0) with noise; the feeder has 400 kW of wind with a first-order autoregressive profile; base loads peak at 90–150 kW per hub with commercial (midday) and residential (evening) shapes (Pfenninger & Staffell, 2016; Staffell & Pfenninger, 2016). The price follows a duck curve (10 ¢/kWh base, an evening peak of 28 ¢/kWh and a midday trough of 5 ¢/kWh on sunny days, scaled by 0.8–1.3 per day with noise) and export is remunerated at half the price; the carbon intensity of imports runs from 300 g/kWh at midday on sunny days to 520 g/kWh at night (Borenstein, 2005; Hirth, 2013; EPA, 2024). A demand charge of 0.6 per kW of the daily peak import (about 18 per kW-month) and a carbon price of 0.05 per kg complete the economics; battery degradation is charged at 0.02 per kWh of throughput, in the range implied by the degradation literature (Xu et al., 2018; Peterson et al., 2010); unmet energy at departure costs 0.6 per kWh; transformer overload is charged at 0.4 per kWh above the rating and under-voltage at 20 per hundredth of a per-unit below the limit per step; and fast reserve is paid at 0.012 per kW of symmetric headroom committed per 15-minute

step (Kempton & Tomić, 2005b; Han et al., 2010).

Figure 4: Testbed example day: (a) base load, PV and wind, (b) price and carbon-intensity profiles, (c) connected EVs per hub, (d) stylised communication archetypes



Communication archetypes

Table 3 gives the three communication archetypes. They are stylised representations, not measurements of any network. Under 4G-LTE, 5% of aggregator messages are lost in a step (the agent keeps the last received message and its age of information grows, capped at eight steps), 20% of setpoints are delayed by one step (the previous action is applied), the aggregator's fleet-flexibility messages (the forced power and the upward and downward flexible capacity of the whole fleet) are not available (only the substation measurement is), and the link does not qualify for fast reserve products. Under 5G, loss is 1%, delay 3%, fleet messages are available and half of the reserve products are eligible. Under 6G, loss is 0.1%, there is no delay, fleet messages are available and all reserve products are eligible. The loss and delay values represent the orders of magnitude discussed in the URLLC and 6G literature (Bennis et al., 2018; Popovski et al., 2019; Saad et al., 2020; ITU, 2023), the age-of-information mechanism follows Kaul et al. and Yates et al. (Kaul et al., 2012; Yates et al., 2021), and reserve eligibility represents the latency and reliability requirements of fast frequency products that only URLLC links meet (Han et al., 2010; Noor-A-Rahim et al., 2022); all four parameters are varied in the degradation sweep and the decomposition experiment.

Table 2: Testbed: feeder and transformer, hubs and chargers, fleets, renewables and base loads, prices, carbon, charges and penalties

Element	Model	Values (stylised)	Basis
Feeder and transformer	Radial feeder, 8 hub buses; substation transformer rating; linearised voltage drop from cumulative downstream flow	Rating 1,200 kW; 0.4% of rated flow per section (×3 at full flow); limit 0.95 pu	Farivar & Low (2013); Baran & Wu (1989); IEEE (2011)
Hubs and chargers	8 hubs, each with a feeder-connection limit and bidirectional chargers	25 EVs per hub (200 total); 11 kW chargers; efficiency 0.93 each way; hub limit 320 kW	ISO (2022); Lee et al. (2019)
Fleets	Workplace fleets at even hubs, residential at odd hubs; uniform arrival and departure windows; arrival SOC; departure target; V2G floor	Capacity 40–80 kWh; arrival SOC 20–60%; target 85%; V2G floor 30%; workplace 07:00–09:30 to 16:00–19:00; residential 15:00–19:30 to 03:00–05:45	Lee et al. (2019); Sadeghianpourhamami et al. (2020)
Renewables	Hub PV with clear-sky shape × daily cloudiness factor + noise; feeder wind with first-order autoregressive profile	PV 150 kW per hub; cloudiness 0.3–1.0; wind 400 kW	Pfenninger & Staffell (2016); Staffell & Pfenninger (2016)
Base loads	Commercial (midday) and residential (evening) shapes with noise	Peak 90–150 kW per hub	Author's stylisation
Prices	Duck curve: base, evening peak, midday trough on sunny days; daily scale and noise; export at a fraction of the price	Base 10, peak 28, trough 5 ¢/kWh; daily scale 0.8–1.3; export at 50% of price	Borenstein (2005); Hirth (2013)
Carbon intensity	Follows solar output: low at midday on sunny days, high at night	300–520 g CO_2 /kWh	EPA (2024); Siler-Evans et al. (2012)
Charges and prices	Demand charge on the daily peak import; carbon price; degradation per kWh of throughput; reserve payment on symmetric headroom	0.6 per kW of daily peak; 0.05 per kg; 0.02 per kWh; 0.012 per kW per step	Kempton & Tomić (2005b); Xu et al. (2018); Peterson et al. (2010); Rennert et al. (2022)
Penalties	Unmet energy at departure; transformer overload per kWh above rating; voltage below limit	0.6 per kWh unmet; 0.4 per kWh overload; 20 per (pu × 100) below 0.95	Author's stylisation; IEEE (2018)
Time	15-minute steps over a 06:00–06:00 day	96 steps; 40 held-out test days; validation days separate from test days	Law (2015)

Environment: observations, safety layer, actions and rewards

Table 4 gives the environment. Each hub agent observes 32 features: the time of day (sine and cosine), its own connected vehicles, energy need, mean laxity, forced power, flexible upward and downward capacity, mean state of charge and vehicles departing within two hours; PV now and one hour ahead; the price now, one and two hours ahead and the mean for the rest of the day; carbon intensity; the aggregator message (net feeder load, fleet forced power, fleet upward and downward capacity) with its age of information; the day's peak import so far; its own base load; and a hub identifier. A safety layer precedes the agent: vehicles whose remaining time at maximum power is just enough to reach their target are charged at the required rate regardless of the action, in the spirit of constrained and safe reinforcement learning for EV charging (Li et al., 2020); the agent's action (one of five levels from −1 to +1) sets the flexible remainder as a fraction of the hub's flexible charging capacity (positive) or discharge capacity (negative), and the hub allocates charging to vehicles by least laxity first and discharging to

vehicles with the highest state of charge and sufficient slack, within the hub's feeder limit. The team reward per step is the negative of the system cost: energy cost at the price (export at half price), the demand charge on any new daily peak (which sums over the day to the charge on the daily maximum), carbon cost, degradation cost, the transformer-overload and voltage penalties, the unmet-energy cost of vehicles departing in the step, less reserve revenue on the symmetric headroom that flexible vehicles retain after dispatch, scaled by the link's eligibility. The individual reward for independent learning is the hub's own energy cost, degradation, unmet energy and reserve, plus its share of the system penalties and demand charge in proportion to its power.

Table 3: Communication archetypes (stylised): message loss, setpoint delay, fleet-flexibility messages, reserve eligibility and the mechanisms they represent

Archetype	Message loss per step	Setpoint delay	Fleet-flexibility messages	Reserve eligibility	What it represents / sources
4G-LTE	5%	20%	No (substation measurement only)	0%	Best-effort cellular links; stale aggregator view; no fast-reserve qualification
5G (URLLC partial)	1%	3%	Yes	50%	URLLC slices for part of the traffic; most reserve products qualify
6G (URLLC + edge)	0.1%	0%	Yes	100%	URLLC with edge intelligence; message age ≈ 0; all fast-reserve products qualify
Mechanisms	Lost message: agent keeps the last received aggregator message; age of information grows by one step (cap 8)	Delayed setpoint: previous action applied for one step	Fleet-flexibility message: forced power, upward and downward flexible capacity of the whole fleet, net feeder load	Share of fast reserve products the link qualifies for; revenue scaled accordingly	Bennis et al. (2018); Popovski et al. (2019); Kaul et al. (2012); Yates et al. (2021); Han et al. (2010); Noor-A-Rahim et al. (2022); ITU (2023)

Learning agents, baselines and the bound

Table 5 gives the agents. Both learning policies use a shared multilayer perceptron (two hidden layers of 64 rectified units) that maps a hub's observation, including its identifier, to action values, trained with Double DQN targets (Mnih et al., 2015) on three-step returns, a replay buffer of 80,000 transitions, minibatches of 64 transitions (512 hub-rows) with two updates per environment step, a target network updated every 500 steps, Adam at 5×10^{-4}, discount 0.98, Huber-clipped errors, and ε-greedy exploration decaying from 1.0 to 0.05 over 300 episodes. Two training-only devices make the small network learn within the budget: a shaping term proportional to the square of the substation import relative to the rating, which gives a dense signal for the daily peak that the demand charge rewards only sparsely, and a constant per-step offset that centres returns near zero; neither is used at evaluation, where policies are compared on the unshaped system cost. Independent Q-learning (IQL) trains each hub's action value on its individual reward (Tan, 1993); the value-decomposition network (VDN) sums the eight hubs' action values into a team value trained on the team reward, which is centralised training with decentralised execution; each hub executes its own greedy action from its own observation (Sunehag et al., 2018; Rashid et al., 2018). Each policy is trained for 500 episodes (days drawn from the

testbed's distributions) with three seeds, and its learning curve records the greedy policy's mean cost on six fixed validation days every 25 episodes; VDN is trained separately under each communication archetype, IQL under 6G. Two baselines and a bound frame the comparison: uncontrolled charging (every flexible vehicle charges at full rate from arrival), a time-of-use heuristic (charge at full rate when the price is in the lowest 35% of the day or PV exceeds base load, hold otherwise, discharge flexible vehicles when the price is in the top 15% and laxity allows, charge at half rate when laxity is low), and a perfect-information linear programme that dispatches an aggregated energy envelope per hub (Sortomme & El-Sharkawi, 2012; Mukherjee & Gupta, 2015) (summed vehicle energy between a lower envelope (the latest trajectory that still meets each departure target) and an upper envelope (the earliest), with exact energy at departure, transformer and hub limits, demand charge, degradation, carbon and reserve) solved per day with HiGHS (Virtanen et al., 2020). The oracle knows every arrival, departure, price and renewable profile in advance and relaxes vehicle-level constraints, so it is an upper bound on achievable performance, not a policy.

Table 4: Environment model: observations, safety layer, action and allocation, power flow and constraints, reward components, individual reward

Element	Specification
Observation (32 features per hub)	Time of day (sine, cosine); own connected vehicles, energy need, mean laxity, forced power, flexible upward and downward capacity, mean SOC, vehicles departing within 2 h; PV now and +1 h; price now, +1 h, +2 h and mean for the rest of the day; carbon intensity; aggregator message (net feeder load, fleet forced power, fleet upward and downward capacity) and its age of information; peak import so far; own base load; hub identifier (one-hot)
Safety layer	Vehicles whose remaining steps at maximum power are just enough to reach the target (one-step margin) are charged at the required rate regardless of the action; the agent controls only the flexible remainder
Action	One of five levels (−1, −0.5, 0, +0.5, +1): fraction of the hub's flexible charging capacity (positive) or discharge capacity (negative); the edge aggregator scales flexible charging down when its (possibly stale) estimate of the substation import would exceed the rating
Allocation within a hub	Forced vehicles at their forced rate; flexible charging by least laxity first; discharge by highest SOC among vehicles with slack (laxity > 0.3) above the V2G floor; within the hub feeder limit
Power flow and constraints	Substation import = base loads + hub power − PV − wind; demand charge on the daily peak; linearised voltage along the feeder; transformer overload and under-voltage penalties
Team reward (VDN)	−(energy cost − export revenue + demand-charge increment + carbon cost + degradation + overload and voltage penalties + unmet-energy cost of departing vehicles − reserve revenue); during training a shaping term $w \cdot (\text{import}/\text{rating})^2$ per step rewards a flat import profile and a constant per-step offset centres returns near zero (neither changes the ranking of policies at evaluation)
Individual reward (IQL)	Own energy cost, degradation, unmet energy and reserve revenue, plus the hub's share of the system penalties, carbon cost and demand-charge increment in proportion to its absolute power
Reserve	Symmetric headroom retained by flexible vehicles after dispatch, min(up, down) × 0.5, paid per kW per step and scaled by the archetype's eligibility

.

Table 5: Learning agents, baselines and the perfect-information bound: architecture, targets, hyperparameters, training protocol, heuristic rules and oracle formulation

Policy	Specification
Shared Q-network	Multilayer perceptron 32→64→64→5 (rectified units), hub identifier as input; Double DQN targets; 3-step returns; replay buffer of 80,000 transitions; minibatches of 64 transitions (512 hub-rows), 2 updates per environment step; target network every 500 steps; Adam at 0.0005; discount 0.98; ε-greedy from 1.0 to 0.05 over 300 episodes; Huber-clipped errors; NumPy implementation with hand-written backpropagation
IQL (independent DQN)	Each hub's action value trained on its individual reward; hubs share parameters but act and learn independently (Tan, 1993)
VDN-MARL	$Q_{tot} = \Sigma_i Q_i(o_i, a_i)$ trained on the team reward; centralised training, decentralised execution; each hub executes the greedy action of its own Q_i (Sunehag et al., 2018; Rashid et al., 2018)
Training protocol	500 episodes (days drawn from the testbed distributions) × 3 seeds; VDN trained separately under 4G, 5G and 6G links, IQL under 6G; learning curve = greedy cost on 6 fixed validation days every 25 episodes
Uncontrolled	Every flexible vehicle charges at full rate from arrival (no aggregator projection)
Time-of-use heuristic	Charge when the price is in the lowest 35% of the day or PV exceeds base load, at a rate limited by the substation headroom reported in the (possibly stale) aggregator message; discharge flexible vehicles when the price is in the top 15% and laxity exceeds 0.5; charge low-laxity hubs at half rate; hold otherwise
LP oracle (bound)	Perfect information on arrivals, departures, prices and renewables; dispatches an aggregated energy envelope per hub between the latest and earliest feasible trajectories with exact energy at departure; transformer rating with priced overload slack, hub limits, demand charge, degradation, carbon and reserve; solved per day with HiGHS (Sortomme & El-Sharkawi, 2012; Mukherjee & Gupta, 2015; Virtanen et al., 2020)

Evaluation, experiments and economics

Table 6 lists the experiments and Table 7 summarises the methods. All policies are evaluated on the same 40 held-out days (common random numbers) under each archetype; learned policies are pooled over three seeds. Outcomes are the daily system cost and its components, the daily peak import, transformer-overload steps and energy, voltage violations, unmet energy at departure, PV self-consumption (PV used on the feeder divided by PV generated), export, emissions, degradation cost, V2G discharge, reserve revenue and the mean age of information; paired differences against uncontrolled charging give the probability of improvement over days . The communication-degradation sweep evaluates the 6G-trained VDN policies under message loss of 0–30% and setpoint delay of 0–50% without retraining; the decomposition experiment evaluates them under 6G links without reserve, without fleet messages, with 4G loss only, with 4G delay only, and under 4G links with full reserve, and compares with VDN retrained on 4G links, which separates the information channel, the actuation channel and the market channel of the 6G effect. The scenario grid evaluates uncontrolled, time-of-use, VDN-4G, VDN-6G and the oracle on 12 days per cell across EV penetration (15, 25, 40 per hub), PV scale (0, 1, 2 × 150 kW per hub) and price volatility (baseline, doubled), without retraining: a generalisation test. The Monte Carlo analysis draws 2,000 sets of economic inputs (price scale (0.7–1.6), demand charge (0.3–1.0 per kW), carbon price (0–0.19 per kg) (Rennert et al., 2022; EPA, 2023), degradation cost (0.015–0.06 per kWh), reserve price (0.004–0.02 per kW-step), unmet-energy cost (0.3–1.0 per kWh) and a connectivity premium of 6G over 4G of 0.5–10 per EV-month) and recomputes the annual cost of each policy from its logged test-day trajectories (the policies are fixed, so this is a valuation of recorded behaviour, not a re-optimisation), reporting the distribution of annual savings per vehicle and the probability that the net value of 6G over 4G is positive; Sobol indices attribute the variance of that net value to the inputs (Sobol', 2001; Saltelli et al., 2010; Herman & Usher, 2017; Iwanaga et al., 2022). A gradient-boosting metamodel on the scenario grid crossed with sampled economics, with a depth-3 tree as a readable summary, says which conditions make the saving per vehicle large

(Ke et al., 2017; Friedman, 2001; Rudin, 2019; Molnar, 2025).

Table 6: Experiments: evaluation protocol, communication-degradation sweep, decomposition of the 6G effect, scenario grid, Monte Carlo economics, Sobol and metamodel

Experiment	Design	Outputs
Evaluation on held-out days	40 test days (common random numbers) × 3 archetypes × 5 policies; learned policies pooled over 3 seeds; paired differences against uncontrolled charging	Daily cost and components, peak, overload, voltage, unmet energy, PV self-consumption, emissions, degradation, V2G energy, reserve revenue, age of information; probability of improvement
Communication-degradation sweep	VDN-6G policies evaluated without retraining under message loss 0–30% and setpoint delay 0–50% (11 settings) on 20 test days	Cost, unmet energy, peak, overload and age of information versus loss and delay
Decomposition of the 6G effect	VDN-6G policies under 6G links without reserve, without fleet messages, with 4G loss only, with 4G delay only, and under 4G links with full reserve; compared with VDN retrained on 4G	Shares of the 6G effect attributable to the market, information and actuation channels
Scenario grid	EV penetration (15, 25, 40 per hub) × PV scale (0, 1, 2 × 150 kW per hub) × price volatility (baseline, doubled): 18 cells × 12 days; uncontrolled, time-of-use, VDN-4G, VDN-6G and oracle without retraining	Generalisation of the trained policies; savings and constraint outcomes by condition
Monte Carlo economics	2,000 draws of price scale, demand charge, carbon price, degradation cost, reserve price, unmet-energy cost and connectivity premium; annual cost of each policy from logged test-day trajectories (policies fixed)	Distributions of annual savings per vehicle; probability that the net value of 6G over 4G is positive
Sobol indices and metamodel	Saltelli sampling over the seven economic inputs; LightGBM on the scenario grid crossed with sampled economics (60 draws per cell), with a depth-3 tree as a readable summary	First-order and total indices of the net value of 6G; drivers of the saving per vehicle

Verification

Verification checks were that the per-step energy balance closes (import minus export equals base load plus hub power minus PV and wind), that the demand-charge increments sum to the charge on the daily peak, that the oracle's linear programme is feasible on every held-out day and its dispatch respects the envelopes, that validation learning curves fall and flatten within the training budget and that trained policies beat uncontrolled charging on held-out days for every seed, that the safety layer holds unmet energy to the irreducible minimum (vehicles whose sessions are too short to reach the target), that policies trained under 6G degrade monotonically as loss and delay rise, and that the oracle's cost lies below every policy's on every day (Sargent, 2013; Law, 2015). The code is organised so that the testbed, the archetypes, the environment and the agents can be replaced independently.

Table 7: Methods: testbed, archetypes, environment, agents, training, evaluation, experiments, verification and software

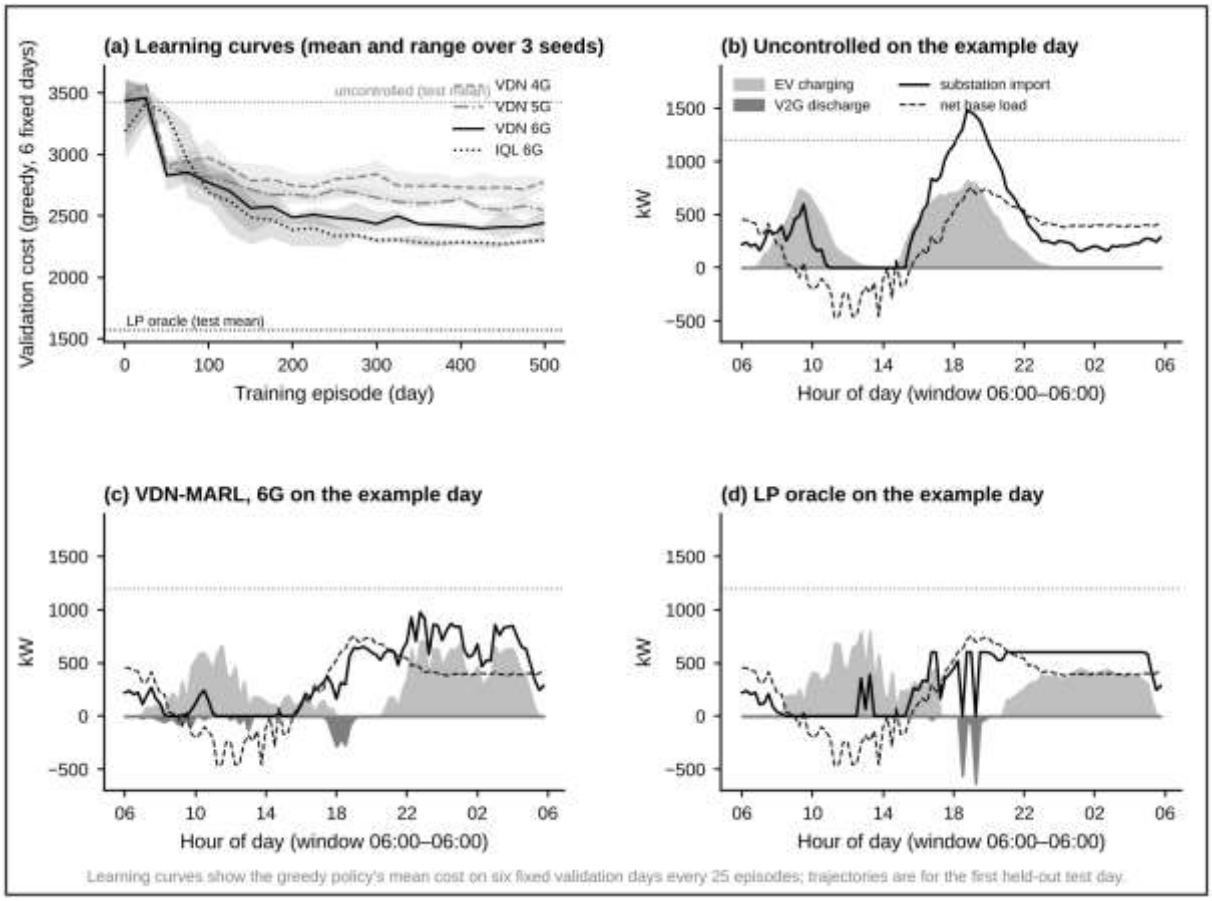
Element	Setting
Design	Design-science research with an artificial evaluation on a simulation testbed (Peffers et al., 2007; Venable et al., 2016)
Testbed	Radial feeder, 8 hubs, 200 EVs, 1,200 kW transformer, hub PV and feeder wind, duck-curve prices, carbon intensity, demand charge, reserve market (Table 2)
Communication	Three stylised archetypes (4G-LTE, 5G, 6G) by loss, delay, fleet messages and reserve eligibility (Table 3); age of information mechanism
Environment	32-feature observations, safety layer, five-level actions, least-laxity allocation, aggregator projection, linearised voltage, team and individual rewards (Table 4)
Agents	IQL and VDN with a shared Double-DQN network, n-step returns, replay and target network; uncontrolled and time-of-use baselines; perfect-information LP oracle (Table 5)
Training	500 episodes × 3 seeds per policy–archetype pair; greedy validation every 25 episodes
Evaluation	40 held-out days with common random numbers; paired comparisons; interval estimates over days and seeds
Experiments	Degradation sweep, decomposition of the 6G effect, scenario grid, Monte Carlo economics, Sobol indices, metamodel (Table 6)
Verification	Energy balance, demand-charge accounting, oracle feasibility and envelope compliance, learning-curve convergence, safety floor on unmet energy, monotone degradation, bound below every policy
Software	Python, NumPy, SciPy (HiGHS), SALib, LightGBM, scikit-learn; results file results18.json and training_curves18.json; runtime 50 minutes on two cores

Results

Learning and policy behaviour

Figure 5 shows the learning curves and three policies on the first held-out day. Validation cost (the greedy policy's mean on six fixed days) fell from 3,200–3,470 at the start of training to 2,775 (VDN under 4G), 2,547 (5G), 2,441 (6G) and 2,302 (IQL under 6G) at episode 500, with the best checkpoints 0.4–4% lower, and the curves had flattened by episode 300–400 in every run; seeds differed by up to about 10% of cost at the end of training, which is the spread the evaluation carries forward. The trajectories show what the policies learned. Uncontrolled charging (Figure 5b) charges the workplace fleet at full rate as it arrives between 07:00 and 09:30, lifting the substation import above the 1,200 kW rating during the morning base-load rise, and charges the residential fleet from the moment it arrives into the evening price peak; the VDN policy under 6G (Figure 5c) delays the morning charge into the midday photovoltaic surplus, holds the residential fleet through the evening peak and charges it overnight at a rate the aggregator's projection keeps under the rating, with some discharge at the price peak; the oracle (Figure 5d) does the same with foresight, charging at a nearly constant rate that leaves the daily peak at the base-load level and discharging briefly at the evening price maximum. The independent learner (IQL) under 6G learned a different policy from VDN: it almost never discharges (3 kWh per day against 693 kWh for VDN) and instead holds the flexible vehicles in the state-of-charge window that earns the reserve payment: a strategy that the individual reward, which credits each hub directly with its own reserve revenue and degradation, makes easy to learn.

Figure 5: Learning and behaviour: (a) validation learning curves of VDN under 4G, 5G and 6G links and of IQL under 6G (mean and range over three seeds), (b) uncontrolled charging, (c) VDN-MARL under 6G and (d) the LP oracle on the first held-out day



Outcomes by policy and network

Table 8 and Figure 6 give the outcomes on the 40 held-out days under each archetype; Figure 7 decomposes the daily cost and gives the paired savings, and Table 9 gives the paired statistics with seeds and the gap to the bound. Uncontrolled charging costs 3,426 per day (5th–95th percentile over days 2,957–4,137), overloads the transformer on 6.5 steps per day with 208 kWh above the rating, violates the voltage limit on 13 steps, and has a peak import of 1,407 kW against a rating of 1,200 kW; its penalty component is 417 per day and its demand charge 844. Every other policy removes the overload. Under 6G links the time-of-use heuristic costs 2,214 (−35.4%, probability of beating uncontrolled charging on a given day 100%), VDN 2,158 (−37.0%), IQL 2,042 (−40.4%) and the oracle 1,573 (−54.1%). The learned policies therefore capture only part of the headroom between the heuristic and the bound under 6G (9% of it for VDN and 27% for IQL): the gap to the bound is 40.8% of the bound's cost for the heuristic, 37.2% for VDN and 29.8% for IQL. The components differ in instructive ways. The heuristic reaches its saving by removing the overload and by discharging 916 kWh per day at the price peak, at a degradation cost of 162; VDN discharges 693 kWh and earns 205 of reserve revenue; IQL discharges almost nothing, earns 286 of reserve revenue, keeps degradation at the uncontrolled level (117) and leaves only 1.9 kWh of unmet energy per day against 11–15 kWh for VDN and the heuristic: the best user service of any policy short of the oracle. Photovoltaic self-consumption rises from 92.6% under uncontrolled charging and 93.8% under the heuristic to 97.9–98.3% for the learned policies and 99.1% for the oracle, and emissions fall from 5.13 t per day to 4.90–4.96 t for the learned policies and 4.67 t for the oracle (the heuristic, which exports more and imports at night, emits slightly more than uncontrolled charging at 5.23 t). Voltage violations fall from 13 steps per day to 3–7 for the learned policies under 6G and none for the oracle.

Two further outcomes matter to the people on either side of the controller. For the users, unmet energy at departure is the service metric: uncontrolled charging never leaves a vehicle short, the oracle never

does by construction, IQL leaves 1.9 kWh per day across 200 vehicles (0.03% of the 5,830 kWh it delivers), VDN 11.3 kWh under 6G and 20.8 under 4G (0.2–0.3% of the energy delivered) and the heuristic 14.5; these shortfalls cost 1–13 per day at the unmet-energy price, a small share of the saving, but they are concentrated on vehicles with short sessions and would be visible to their drivers. For the utility, the daily peak is the asset metric: uncontrolled charging peaks at 1,274–1,580 kW across the 40 days (5th–95th percentile), above the rating on most of them; the heuristic at 1,007–1,178; VDN under 6G at 917–1,243, with the upper tail touching the rating on the days with the largest residential fleets; IQL at 905–1,200; and the oracle at 522–686, which shows how much of the peak is flexible when departures are known in advance. The energy imported from the grid falls from 10,063 kWh per day under uncontrolled charging to 9,582–9,720 for the learned policies and 9,296 for the oracle, and export falls from 905 to 425–450 kWh, because the learned policies absorb the midday photovoltaic surplus into the workplace fleet; the heuristic, which charges whenever the price is low or photovoltaics exceed base load but discharges 916 kWh at the evening peak, imports slightly more than uncontrolled charging and exports as much.

Table 8: Outcomes on 40 held-out days by policy and network: daily cost, energy cost, peak and demand charge, emissions, degradation, unmet energy, overload and voltage steps, PV self-consumption, V2G energy, reserve revenue and age of information

Policy, network	Daily cost, mean (5th–95th over days)	Energy cost	Peak kW / demand charge	CO_2 (kg)	Degradation	Unmet kWh	Overload / voltage steps	PV self-use (%)	V2G kWh / reserve rev.	Mean AoI (steps)
Uncontrolled, 4G-LTE	3,426 (2,957–4,137)	1,791	1,407 / 844	5,134	116.5	0.0	6.5 / 13.0	93	0 / 0	0.05
Time-of-use heuristic, 4G-LTE	2,358 (2,022–2,828)	1,230	1,085 / 651	5,206	162.0	12.7	0.0 / 9.2	94	927 / 0	0.05
VDN-MARL, 4G-LTE	2,497 (2,049–3,115)	1,271	1,126 / 676	5,061	176.1	20.8	0.3 / 9.9	98	1,091 / 0	0.05
LP oracle (bound), 4G-LTE	1,960 (1,623–2,388)	1,161	684 / 410	4,712	152.7	0.0	0.0 / 0.0	99	838 / 0	0.00
Uncontrolled, 5G	3,426 (2,957–4,137)	1,791	1,407 / 844	5,134	116.5	0.0	6.5 / 13.0	93	0 / 0	0.01
Time-of-use heuristic, 5G	2,288 (1,915–2,761)	1,236	1,082 / 649	5,224	161.8	14.0	0.0 / 8.8	94	916 / 74	0.01
VDN-MARL, 5G	2,289 (1,876–2,836)	1,279	1,080 / 648	4,994	163.8	15.8	0.1 / 4.6	99	926 / 95	0.01
LP oracle (bound), 5G	1,770 (1,422–2,210)	1,211	636 / 382	4,685	137.1	0.0	0.0 / 0.0	99	478 / 194	0.00
Uncontrolled, 6G	3,426 (2,957–4,137)	1,791	1,407 / 844	5,134	116.5	0.0	6.5 / 13.0	93	0 / 0	0.00
Time-of-use heuristic, 6G	2,214 (1,837–2,673)	1,237	1,082 / 649	5,228	161.9	14.5	0.0 / 8.8	94	916 / 147	0.00
IQL (independent DQN), 6G	2,042 (1,625–2,688)	1,320	1,016 / 610	4,902	116.8	1.9	0.1 / 6.8	98	3 / 286	0.00
VDN-MARL, 6G	2,158 (1,744–2,657)	1,298	1,065 / 639	4,960	150.6	11.3	0.2 / 3.3	98	693 / 205	0.00
LP oracle (bound), 6G	1,573 (1,218–2,024)	1,260	598 / 359	4,667	124.2	0.0	0.0 / 0.0	99	179 / 403	0.00

Figure 6: Outcomes by policy and network on the held-out days: (a) daily cost, (b) peak substation import, (c) unmet energy at departure, (d) PV self-consumption

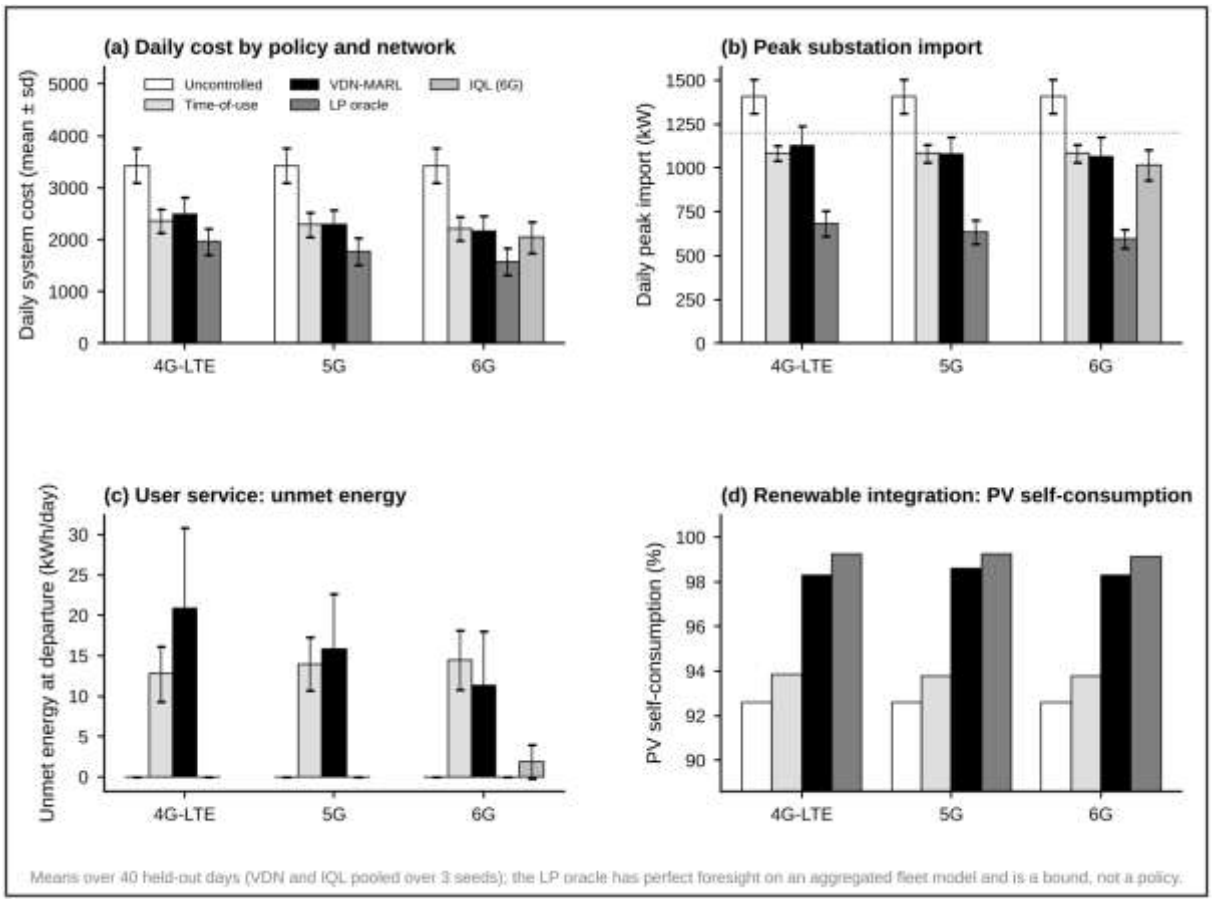


Figure 7: (a) Decomposition of the daily cost into energy, demand charge, degradation, unmet energy, penalties, carbon and reserve revenue for each policy; (b) paired savings against uncontrolled charging with the probability of improvement

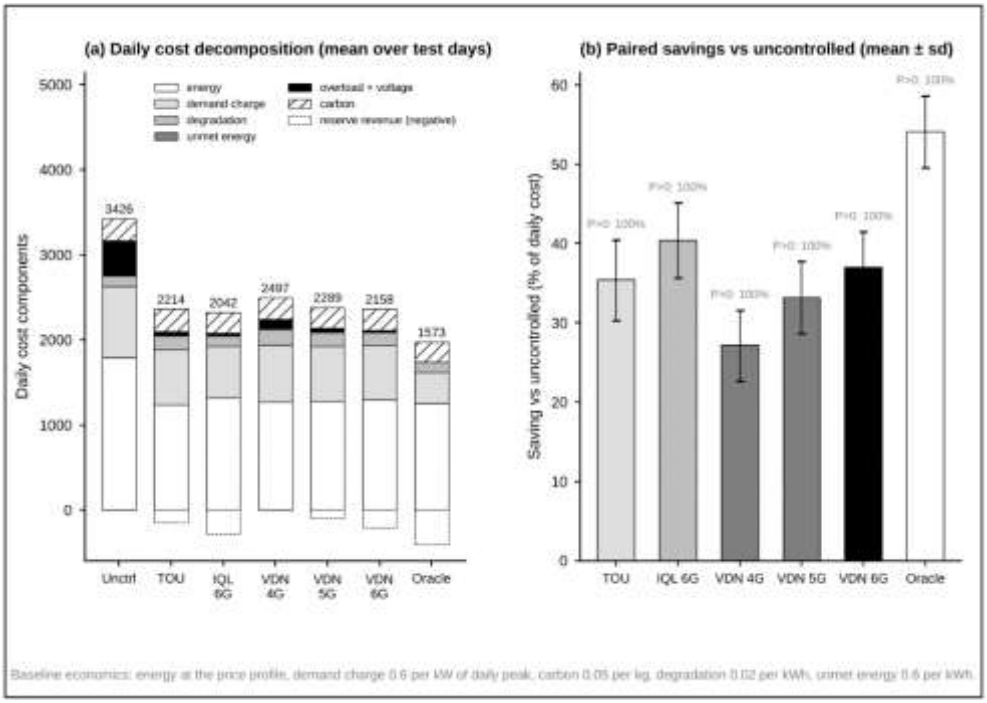


Table 9: Paired comparison with uncontrolled charging on the held-out days: mean saving, dispersion over days, probability of improvement, results by seed, gap to the oracle and annual saving per vehicle at baseline economics

Policy, network	Mean daily saving vs uncontrolled (share of its cost)	SD over days	P(better than uncontrolled), %	Mean cost by seed	Gap to oracle (%)	Annual saving per EV (baseline economics)
Time-of-use, 6G	1,212 (35.4%)	174	100	n/a	40.8	2,212
IQL, 6G	1,385 (40.4%)	162	100	2,025 / 2,066 / 2,034	29.8	2,527
VDN, 4G-LTE	930 (27.1%)	153	100	2,418 / 2,522 / 2,550	27.4	1,697
VDN, 5G	1,137 (33.2%)	155	100	2,221 / 2,318 / 2,327	29.3	2,076
VDN, 6G	1,268 (37.0%)	154	100	2,066 / 2,277 / 2,132	37.2	2,314
LP oracle, 6G	1,854 (54.1%)	155	100	n/a	0	3,383

The ordering of the learned policies across networks is the article's central result. Under 4G links VDN costs 2,497 per day, 6% more than the heuristic's 2,358: without fleet-flexibility messages, with 5% message loss and 20% delayed setpoints, and with no reserve eligibility, the learned policy over-discharges (1,091 kWh per day), leaves 20.8 kWh of unmet energy and pays 108 in penalties, and a capacity-aware heuristic does better. Under 5G the two are level (2,289 against 2,288). Under 6G the learned policy is 2.5% better than the heuristic and IQL 7.8% better. The value of the network to the learned policy is thus 339 per day between 4G and 6G (27.1% to 37.0% of the uncontrolled cost), against 144 per day for the heuristic (which gains only the reserve revenue the 6G link makes eligible) and 387 per day for the oracle. Across three seeds the VDN-6G policies cost 2,066, 2,277 and 2,132 and the IQL policies 2,025, 2,066 and 2,034: two of the three VDN seeds and all three IQL seeds beat the heuristic under 6G (on 39, 9 and 32 of the 40 days for VDN and 39, 34 and 37 for IQL), and no VDN seed did under 4G (2,418–2,550 against 2,358, winning on at most 13 days).

The value of connectivity: degradation and decomposition

Table 10 and Figure 8(a, b) give the degradation sweep. Evaluated without retraining under rising message loss, the 6G-trained VDN policies' cost rose from 2,184 (no loss, no delay, 20-day subset) to 2,193 at 5% loss, 2,217 at 20% and 2,246 at 30% (+2.8%), with the mean age of information rising to 0.42 steps, unmet energy from 10.8 to 15.1 kWh per day, the peak from 1,066 to 1,105 kW and overload steps from 0.17 to 0.43 per day; the degradation is monotone in loss, as the verification checks require, but small; the policies are robust to stale aggregator messages because the hub's own observation carries most of what it needs. Setpoint delay alone had no measurable effect (2,185–2,189 at 10–50% delay), because a one-step delay of a hub-level fraction changes little when the fraction itself changes slowly. Table 11 and Figure 8(c, d) decompose the 6G effect by evaluating the 6G-trained policies under modified links. Removing reserve eligibility alone raises the cost by 205 per day (the reserve revenue, and 61% of the 338 difference between the 6G-trained and 4G-trained policies); removing the fleet-flexibility messages alone raises it by 220 (65%), with the peak rising to 1,103 kW and penalties to 94 per day, because a policy trained to read the fleet's forced and flexible power misjudges the headroom when those features are frozen; adding 4G loss alone (+6) or 4G delay alone (0) does almost nothing; and 4G links with full reserve eligibility cost the same as 6G links without fleet messages (2,379). The channels are not additive (the market and information channels together exceed the whole difference) because the policy retrained under 4G recovers part of the loss by learning not to depend on the messages. Figure 8(d) shows where the network changes the cost for the three VDN policies: reserve revenue (0, 95, 205 per day under 4G, 5G, 6G), degradation (176, 164, 151), the demand charge (676, 648, 639), and penalties (108, 34, 21). The actuation channel (loss and delay at the 4G levels) is worth little in this testbed; the market channel and the information channel are worth 200–220 per day each, and most of the information channel's value is the ability to use the fleet-flexibility message the edge aggregator provides.

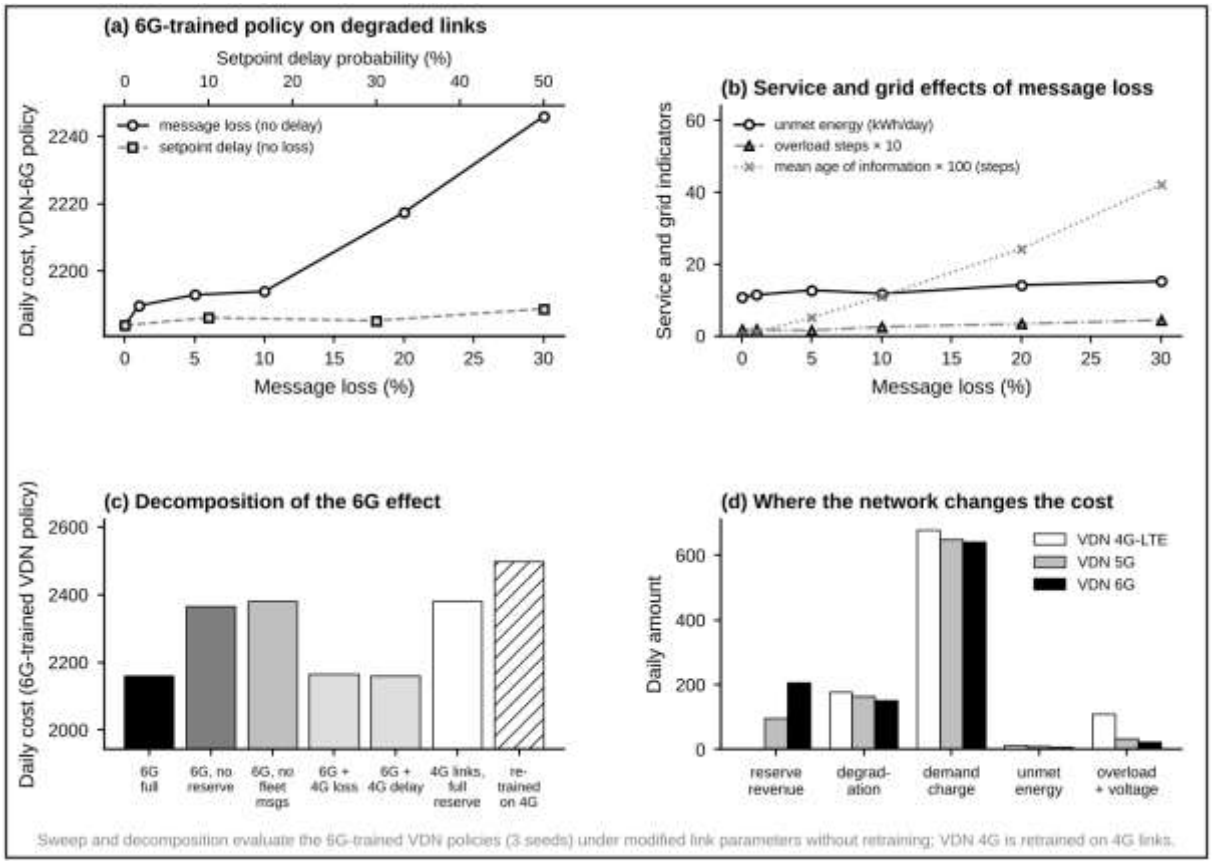
Table 10: Communication-degradation sweep: cost, unmet energy, peak, overload steps and age of information of the 6G-trained VDN policies under message loss of 0–30% and setpoint delay of 0–50%

Message loss	Setpoint delay	Daily cost, mean (SD)	Change vs no loss, no delay (%)	Unmet kWh	Peak kW	Overload steps	Mean AoI (steps)
0%	0%	2,184 (282)	+0.0	10.8	1,066	0.2	0.00
1%	0%	2,190 (286)	+0.3	11.4	1,071	0.2	0.01
5%	0%	2,193 (287)	+0.4	12.7	1,073	0.1	0.05
10%	0%	2,194 (289)	+0.5	11.7	1,069	0.2	0.11
20%	0%	2,217 (284)	+1.5	14.1	1,086	0.3	0.24
30%	0%	2,246 (311)	+2.8	15.1	1,105	0.4	0.42
0%	10%	2,186 (279)	+0.1	11.2	1,068	0.2	0.00
0%	30%	2,185 (287)	+0.1	9.7	1,063	0.2	0.00
0%	50%	2,189 (283)	+0.2	10.2	1,067	0.1	0.00
10%	20%	2,200 (289)	+0.7	10.4	1,075	0.2	0.11
20%	40%	2,211 (289)	+1.2	9.6	1,077	0.2	0.24

Table 11: Decomposition of the 6G effect: 6G-trained VDN policies under 6G links without reserve, without fleet messages, with 4G loss only, with 4G delay only, under 4G links with full reserve, and VDN retrained on 4G

Condition (VDN-6G policies unless stated)	Daily cost	Change vs 6G reference	Share of the 6G–4G difference (%)	Reserve revenue	Unmet kWh	Peak kW
VDN trained and evaluated under 6G (reference)	2,158	0	n/a	205.5	11.3	1,065
6G links, no reserve eligibility (market channel removed)	2,364	+205	61	0.0	11.3	1,065
6G links, no fleet-flexibility messages (information channel removed)	2,378	+220	65	178.3	10.0	1,103
6G links with 4G message loss (5%)	2,164	+6	2	205.5	12.9	1,068
6G links with 4G setpoint delay (20%)	2,158	+0	0	204.4	10.6	1,064
4G links (loss, delay, no fleet messages) with full reserve eligibility	2,379	+221	65	178.1	10.2	1,105
VDN retrained and evaluated under 4G-LTE	2,497	+338	100	0.0	20.8	1,126

Figure 8: The value of connectivity: (a) cost of the 6G-trained VDN policy under message loss and setpoint delay, (b) service and grid effects of message loss, (c) decomposition of the 6G effect into market, information and actuation channels, (d) cost components by network



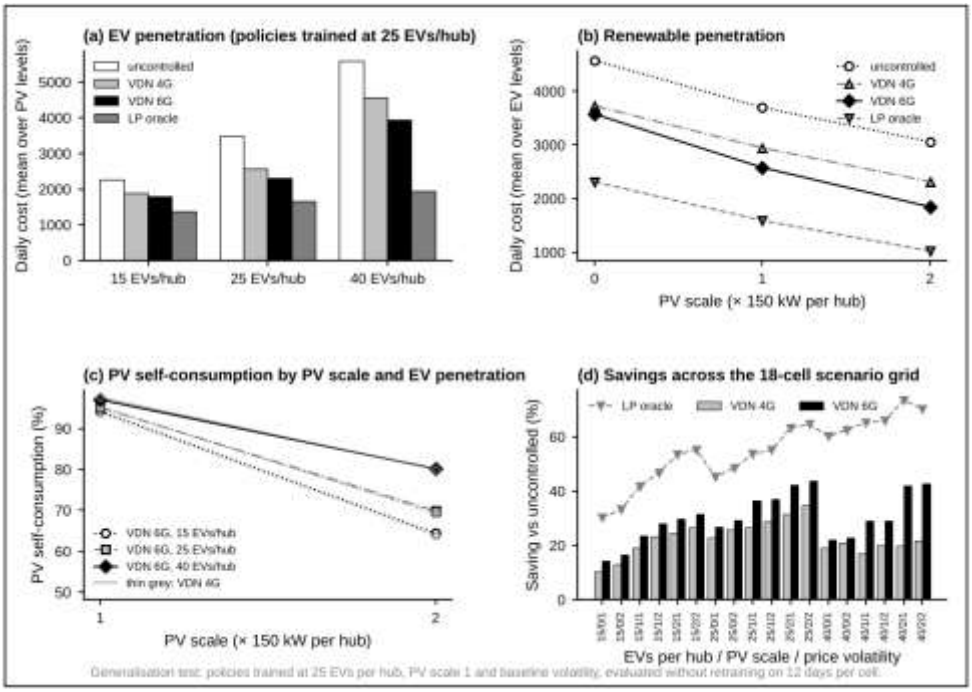
Generalisation across penetration, renewables and prices

Table 12 and Figure 9 give the 18-cell scenario grid, on which the policies trained at 25 vehicles per hub, one unit of photovoltaics and baseline volatility were evaluated without retraining. Three patterns stand out. First, with photovoltaics present the VDN-6G policy beat the heuristic in ten of the twelve cells (by 9–13% at 15 vehicles per hub, 1–4% at 25 and 4–20% at 40) and lost to it by about 2% in two cells with doubled volatility; the saving against uncontrolled charging rose with photovoltaic scale (at 25 vehicles per hub and baseline volatility: 27% at no photovoltaics, 36% at one unit, 42% at two) and with price volatility, because midday surplus and a sharper evening peak are what a flexible fleet can exploit. Second, without photovoltaics the learned policy generalised poorly: at 25 and 40 vehicles per hub the heuristic cost 2,917 and 4,405 against 3,077 and 5,232 for VDN-6G, and the VDN-4G policy was worse still; a policy that learned to wait for a midday surplus waits in vain when there is none, which is the known brittleness of learned controllers outside their training distribution and an argument for retraining or for a heuristic fallback when conditions change. Third, at 40 vehicles per hub (a fleet 60% larger than the training fleet) every policy except the oracle overloaded the transformer (VDN-6G on 5–16 steps per day against 12–24 for uncontrolled charging), and the oracle's advantage widened to 60–73% of the uncontrolled cost: the testbed at that penetration needs a transformer upgrade or a tighter controller, and the framework reports both. Photovoltaic self-consumption under VDN-6G fell from 94–98% at one unit of photovoltaics to 64–86% at two, rising with the number of vehicles (64% at 15 per hub, 80–86% at 40 at two units), which quantifies how fleet size sets the renewable share a feeder can absorb.

Table 12: Scenario grid: daily cost of uncontrolled charging, the time-of-use heuristic, VDN-4G, VDN-6G and the oracle, VDN-6G savings, unmet energy, overload steps and PV self-consumption across EV penetration, PV scale and price volatility (12 days per cell, policies not retrained)

EVs per hub	PV scale	Price volatility	Uncontrolled	Time-of-use	VDN-4G	VDN-6G	Oracle	VDN-6G saving vs uncontrolled (%)	6G over 4G (% of uncontrolled cost)	VDN-6G unmet kWh / overload steps	VDN-6G PV self-use (%)
15	0	baseline	2,815	2,436	2,525	2,413	1,966	14.3	4.0	6 / 0.0	0
15	0	doubled	3,161	2,608	2,751	2,639	2,113	16.5	3.5	4 / 0.0	0
15	1	baseline	2,230	1,866	1,808	1,703	1,297	23.7	4.7	9 / 0.0	94
15	1	doubled	2,316	1,846	1,774	1,669	1,236	28.0	4.6	9 / 0.0	94
15	2	baseline	1,732	1,393	1,305	1,218	807	29.7	5.0	7 / 0.0	64
15	2	doubled	1,864	1,422	1,367	1,278	834	31.4	4.8	8 / 0.0	65
25	0	baseline	4,201	2,917	3,243	3,077	2,301	26.7	3.9	7 / 0.5	0
25	0	doubled	4,382	2,890	3,253	3,096	2,258	29.4	3.6	6 / 0.1	0
25	1	baseline	3,441	2,227	2,532	2,189	1,595	36.4	10.0	12 / 0.2	95
25	1	doubled	3,676	2,276	2,617	2,316	1,651	37.0	8.2	14 / 0.0	97
25	2	baseline	2,797	1,687	1,914	1,619	1,027	42.1	10.6	9 / 0.0	70
25	2	doubled	2,984	1,694	1,950	1,675	1,057	43.9	9.2	9 / 0.1	70
40	0	baseline	6,702	4,405	5,419	5,232	2,672	21.9	2.8	11 / 14.2	0
40	0	doubled	7,090	4,504	5,596	5,480	2,655	22.7	1.6	11 / 15.6	0
40	1	baseline	5,419	4,002	4,497	3,843	1,882	29.1	12.1	15 / 7.0	97
40	1	doubled	5,699	3,949	4,559	4,038	1,945	29.1	9.1	13 / 7.2	98
40	2	baseline	4,649	3,386	3,725	2,702	1,238	41.9	22.0	19 / 5.1	80
40	2	doubled	5,502	3,742	4,299	3,146	1,636	42.8	21.0	22 / 5.2	85

Figure 9: Generalisation across the scenario grid: (a) cost by EV penetration, (b) cost by PV scale, (c) PV self-consumption by PV scale and penetration, (d) savings against uncontrolled charging in all 18 cells



Economics under uncertainty

Table 13 and Figure 10 give the Monte Carlo economics, which revalue the logged test-day trajectories of fixed policies under 2,000 draws of prices, charges and costs. At baseline-year economics the VDN-6G policy saves 1,824 per vehicle per year against uncontrolled charging (5th–95th percentile 1,336–2,293), the VDN-4G policy 1,211 (828–1,613) and the heuristic 1,583 (1,165–1,992); the gross value of 6G over 4G connectivity to the learned policy is 613 per vehicle-year (312–911), and after a connectivity premium of 0.5–10 per vehicle-month the net value is 552 (246–859), positive in 100% of draws. Read as a break-even, the gross value of 613 per vehicle-year is a premium of 51 per vehicle-month, or 26 at the 5th percentile, against the 1.5–10 per vehicle-month assumed for the difference between 6G and 4G service; the margin is wide on the testbed, and it would close only if the reserve market that supplies most of the value were unavailable. These are large numbers because the testbed's uncontrolled baseline overloads a transformer and the feeder has a strong duck curve; the ranking, not the magnitude, is the finding. Sobol indices attribute 87% of the variance of the net value of 6G to the reserve price, 6% to the degradation cost and 3% to the connectivity premium, with price scale, demand charge, carbon price and unmet-energy cost contributing 2% or less each: in this testbed the decision to buy 6G connectivity for V2G coordination is a decision about the reserve market. Figure 11 adds the emissions, V2G activity, reserve and degradation comparison and the metamodel. The metamodel of the saving per vehicle across the scenario grid crossed with sampled economics (Table 14, which also records the training statistics of the learned policies) attributes 34% of its variance to photovoltaic scale, 23% to price scale, 18% to price volatility, 17% to vehicles per hub and 8% to the demand charge; its depth-3 tree reads: with photovoltaics, high volatility and prices above baseline the saving is about 1,990 per vehicle-year, with photovoltaics and baseline volatility 1,330–1,640, and without photovoltaics at high penetration and low prices 940; the conditions under which a fleet operator should expect the most and the least from learned coordination.

Table 13: Monte Carlo economics over 2,000 draws: annual savings per vehicle of the heuristic, VDN-4G and VDN-6G against uncontrolled charging, gross and net value of 6G over 4G, input ranges and Sobol indices

Quantity (per EV per year, baseline-year economics)	Mean	Median	5th–95th percentile	P(> 0), %
Time-of-use vs uncontrolled	1,583	1,576	1,165–1,992	100
VDN-4G vs uncontrolled	1,211	1,202	828–1,613	100
VDN-6G vs uncontrolled	1,824	1,831	1,336–2,293	100
VDN-6G vs VDN-4G (gross value of connectivity)	613	616	312–911	100
Net value of 6G over 4G after the connectivity premium	552	559	246–859	100
Inputs (ranges)	Price scale 0.7–1.6 (mode 1.0); demand charge 0.3–1.0 per kW; carbon price 0–0.19 per kg (mode 0.05); degradation 0.015–0.06 per kWh	Reserve 0.004–0.02 per kW-step; unmet 0.3–1.0 per kWh	Connectivity premium: 6G 2–12 minus 4G 0.5–3 per EV-month	n = 2,000
Sobol indices of the net value of 6G (S1 / ST)	reserve price: 0.87 / 0.87; degradation cost: 0.06 / 0.06; 6G premium: 0.03 / 0.03; price scale: 0.02 / 0.02	demand charge: 0.02 / 0.02; carbon price: 0.00 / 0.00; unmet-energy cost: 0.00 / 0.00	2,304 evaluations; mean net value 544 per EV-year	100

Figure 10: Economics under uncertainty: (a) annual savings per vehicle against uncontrolled charging, (b) gross and net value of 6G over 4G connectivity, (c) Sobol indices of the net value of 6G, (d) summary with the probability of positive value

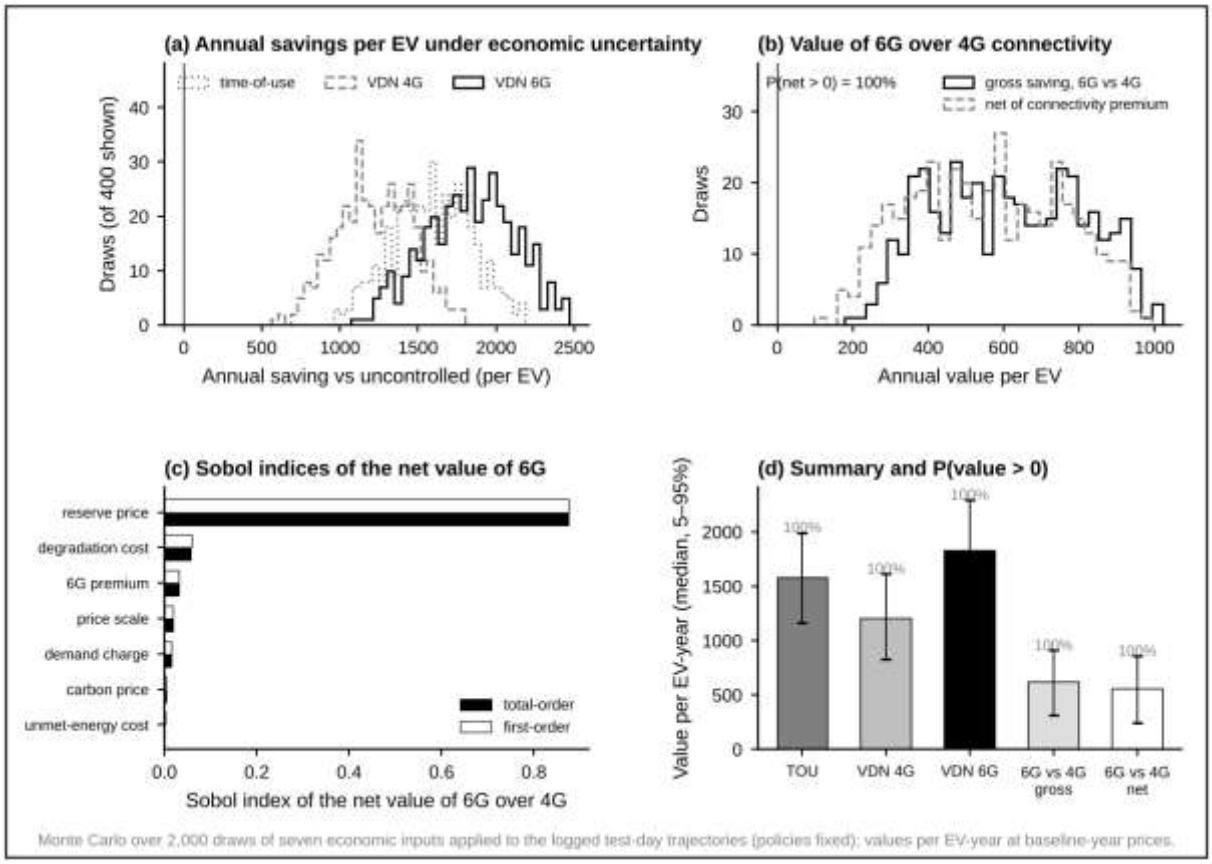


Figure 11: (a) Daily emissions by policy, (b) V2G discharge and feeder export, (c) reserve revenue and battery degradation, (d) metamodel importance of the drivers of the saving per vehicle

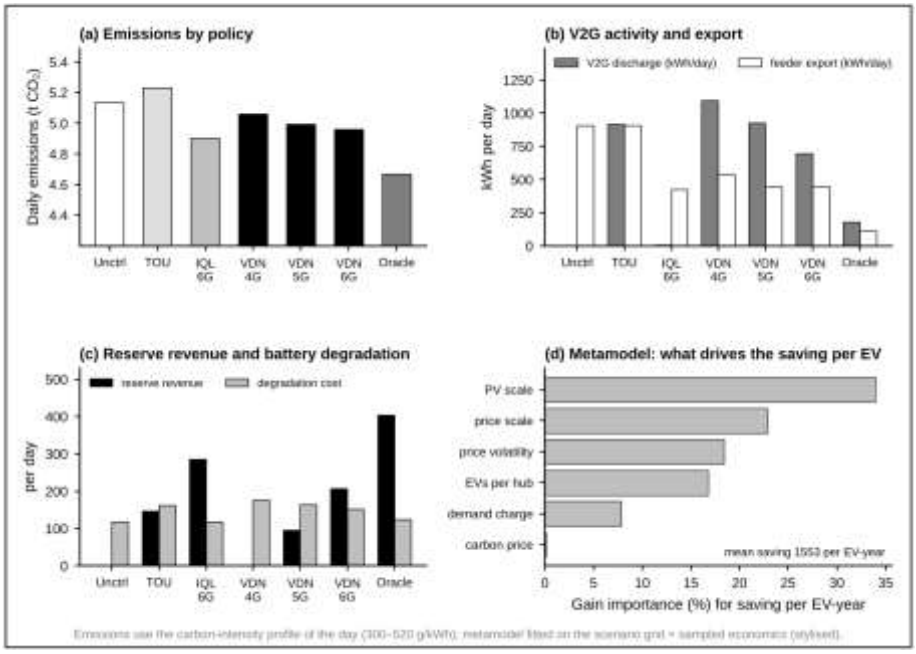


Table 14: Metamodel of the saving per vehicle (importance and depth-3 rules) and training statistics of the learned policies (validation cost by stage, seeds, training time)

Item	Statistics	Detail
Metamodel (LightGBM)	n = 1,080; training R^2 = 1.00; mean saving 1,553 per EV-year	Gain importance: PV scale 34%; price scale 23%; price volatility 18%; EVs per hub 17%; demand charge 8%; carbon price 0%
Depth-3 tree (rules)	Leaf values = saving per EV-year	PV scale ≤ 0.50 & EVs per hub ≤ 32.50 & price volatility ≤ 1.50 → 1,204; PV scale ≤ 0.50 & EVs per hub ≤ 32.50 & price volatility > 1.50 → 1,599; PV scale ≤ 0.50 & EVs per hub > 32.50 & price scale ≤ 1.17 → 936; PV scale ≤ 0.50 & EVs per hub > 32.50 & price scale > 1.17 → 1,176; PV scale > 0.50 & price volatility ≤ 1.50 & EVs per hub ≤ 32.50 → 1,635; PV scale > 0.50 & price volatility ≤ 1.50 & EVs per hub > 32.50 → 1,332; PV scale > 0.50 & price volatility > 1.50 & price scale ≤ 1.02 → 1,617; PV scale > 0.50 & price volatility > 1.50 & price scale > 1.02 → 1,987
VDN, 4G-LTE training	Validation cost: first log 3,468 → last 2,775 (best 2,656); 500 episodes × 3 seeds	Training time 6 min / 7 min / 6 min per seed
VDN, 5G training	Validation cost: first log 3,458 → last 2,547 (best 2,536); 500 episodes × 3 seeds	Training time 7 min / 6 min / 6 min per seed
VDN, 6G training	Validation cost: first log 3,432 → last 2,441 (best 2,366); 500 episodes × 3 seeds	Training time 6 min / 6 min / 6 min per seed
IQL, 6G training	Validation cost: first log 3,186 → last 2,302 (best 2,256); 500 episodes × 3 seeds	Training time 6 min / 6 min / 6 min per seed

Verification

The energy balance closed on every step of every evaluated day; the demand-charge increments summed to the charge on the daily peak; the oracle's linear programme was feasible on all 40 held-out days and all 216 grid days and its dispatch respected the envelopes; the validation curves fell and flattened within the budget; every seed of every learned policy beat uncontrolled charging on every held-out day; unmet energy under the learned policies (1.9–20.8 kWh per day) stayed within the range the safety layer permits (the oracle, with relaxed vehicle constraints, reaches zero); the 6G-trained policies degraded monotonically with message loss; and the oracle's cost lay below every policy's on every day. The run took 50 minutes on two cores, of which training was 12 × 6 minutes.

Discussion

Answers to the research questions

RQ1 asked what a decision framework for connected V2G control requires. The literature yields six requirements (binding constraints, communication as a modelled layer, learning with safety and coordination, evaluation against a bound with seeds and held-out days, decision outputs as probabilities and attributions, and feedback with governance) and the four-layer framework of Figure 1 meets them; the testbed shows the outputs each requirement produces. RQ2 asked how the policies compare. On the testbed every coordinated policy removed the transformer overload that uncontrolled charging causes; the learned policies beat a capacity-aware heuristic only when the network gave them fleet-flexibility messages and reserve eligibility (level under 5G, 2.5–7.8% better under 6G, 6% worse under 4G); the independent learner with individual rewards did better than value decomposition at this budget by learning a low-degradation, reserve-earning policy with the best user service; and the

perfect-information bound stayed 30–41% below the learned and heuristic policies, which is the headroom a better learner would pursue. RQ3 asked what the communication layer is worth. The value of 6G over 4G to the learned policy was 339 per day on the testbed (613 per vehicle-year gross, 552 net after the premium, positive in all economic draws), decomposed into a market channel (reserve eligibility, 205 per day) and an information channel (fleet-flexibility messages, 220 per day) that overlap, with the actuation channel (loss and delay at 4G levels) worth little; the policies were robust to message loss up to 30% (+2.8% cost); and the reserve price explained 87% of the variance of the net value. RQ4 asked how the results generalise. Savings rose with photovoltaic scale and price volatility and fell at high penetration, where the transformer binds for every policy but the oracle; the learned policy failed to generalise to a feeder without photovoltaics; and the metamodel's rules say where to expect 940 and where 1,990 per vehicle-year.

What managed charging alone achieves, and what learning adds

The savings against uncontrolled charging are large (27–40% of daily cost) and most of that is not learning. The uncontrolled baseline overloads the transformer because the morning workplace arrivals charge together on top of the base-load rise, and any managed-charging architecture with the aggregator's grid-safe projection removes the overload and its 417 per day of penalties; the capacity-aware heuristic, which uses the same projection and the same fleet messages, reaches 35% on its own. The increment that learning adds is 2.5% (VDN) to 7.8% (IQL) of the uncontrolled cost under 6G, nothing under 5G and a loss under 4G, and it comes from three behaviours the heuristic does not have: holding vehicles in the state-of-charge window that earns reserve revenue instead of discharging them at the price peak, timing the workplace charge to the photovoltaic surplus rather than to a price threshold, and spreading the overnight residential charge so that the peak stays near the base-load level. An operator reading the framework's outputs should therefore separate the two decisions: whether to manage charging at all, which the testbed answers with the overload and the demand charge, and whether to learn the management policy, which the testbed answers with the network-dependent increment over the heuristic and the much larger gap to the bound. The second decision is the one that depends on connectivity.

Why independent learning did as well as it did

The result that IQL matched or beat VDN deserves comment, because value decomposition is the method the MARL literature recommends for cooperative tasks (Sunehag et al., 2018; Rashid et al., 2018). Three features of the testbed explain it. The aggregator's grid-safe projection handles the one hard coupling constraint (the transformer rating) at execution time, so the hubs do not have to learn to coordinate their charging to avoid overload; the individual reward gives each hub a dense, low-variance signal for the two components that matter most under 6G, reserve revenue and degradation, whereas the team reward spreads credit for them over eight hubs; and the demand charge, the one component that needs coordination, is shared pro rata in the individual reward. Independent learning is known to be a strong baseline when the coupling is weak or handled elsewhere (Tan, 1993; Hernández-Leal et al., 2019; Oroojlooy & Hajinezhad, 2023), and the finding is a reminder that the architecture of the controller (where constraints are enforced and how rewards are attributed) can matter more than the choice of learning algorithm. It is also a result at a particular budget: with 500 episodes and small networks, the harder credit-assignment problem of VDN may simply need more training, and the framework reports the bound so that the reader can see how much either learner left.

What the decomposition says about the communication layer

The decomposition separates three things that the 6G literature tends to present together. The link's loss and delay (the quantities that URLLC targets improve by orders of magnitude) were worth little to this controller at the 4G levels: 5% loss and 20% delayed setpoints on 15-minute hub-level fractions cost 6 per day and nothing respectively, and the sweep found only 2.8% at 30% loss. What was worth 200–220 per day each was the market eligibility that the link's reliability confers and the information that the edge aggregator computes and distributes (the fleet's forced and flexible power) which is a function of the edge layer rather than of the radio. The conclusion for the design of a 6G-enabled V2G system is that the edge aggregator's messages and the reserve qualification are the deliverables, and the link's reliability is their precondition: a 4G link with an edge aggregator and reserve eligibility would, on this testbed, deliver most of the 6G value (the 4G-links-with-full-reserve condition cost 2,379,

between the 6G and 4G policies), whereas a 6G link without them would deliver little. This is consistent with the age-of-information view that what matters is the freshness of the information the controller uses, not the latency of the channel as such (Kaul et al., 2012; Yates et al., 2021), and with the V2X roadmap's emphasis on edge intelligence as the enabler of 6G applications (Noor-A-Rahim et al., 2022; Saad et al., 2020). It is also a caution: a controller that depends on edge messages is only as available as the edge, and the sweep shows that the policy trained with fresh messages degrades gracefully when they are stale but (in the decomposition) sharply when they are absent.

Implications for operators, utilities and aggregators

Three implications follow for practice. First, the decision about connectivity is a decision about markets as much as about control. In the testbed the largest single component of the 6G effect is reserve eligibility (the market channel) and that channel is a property of how system operators qualify flexibility products, not of the radio interface alone; an operator considering a connectivity upgrade should establish which products its fleet could qualify for under each link and at what price before buying the link, in the manner of Kempton and Tomić's original argument that reserves, not energy, repay V2G (Kempton & Tomić, 2005b; Han et al., 2010). Second, the information channel (fleet-flexibility messages from an edge aggregator) is worth as much as the market channel to a learned policy, and a policy trained with that information misbehaves without it; a fleet operator who deploys a learned controller should therefore budget for the monitoring of message age and for a fallback (the time-of-use heuristic in the roadmap of Figure 12) when the link degrades, as the age-of-information literature suggests (Kaul et al., 2012; Yates et al., 2021). Third, the gap between the learned policies and the perfect-information bound is the largest number in the study, and it is a statement about the learner and the budget, not about 6G: a better learner (QMIX, MAPPO, larger networks, longer training, curricula from the heuristic) would move the learned policies toward the bound and raise the value of every channel (Rashid et al., 2018; Yu et al., 2022; Dulac-Arnold et al., 2021), which is why the framework reports the bound at all; it tells the decision-maker how much of the prize the current controller leaves on the table.

Implications for standards, regulation and governance

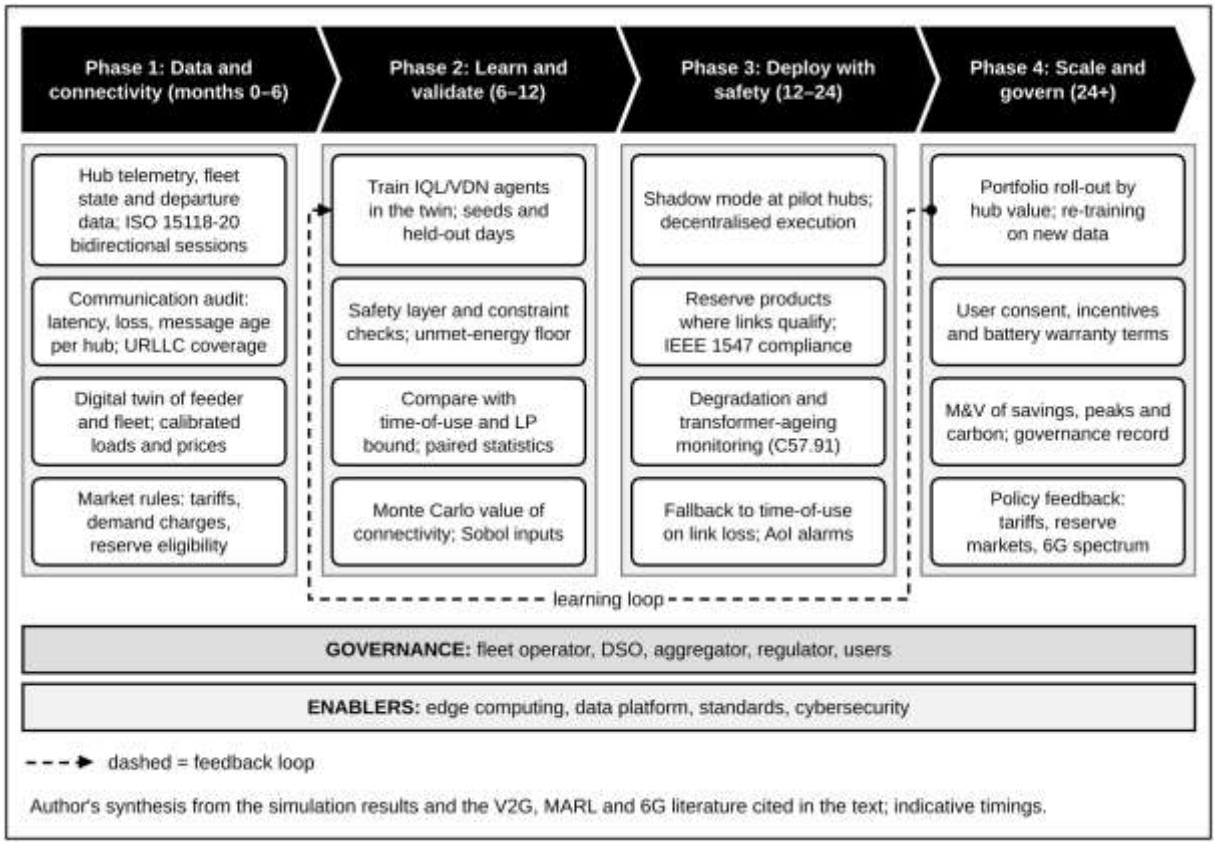
The results connect to four standards and regulatory threads. ISO 15118-20 makes bidirectional sessions and scheduling negotiable between vehicle and charger, which is the interface through which the hub-level setpoints of this study would reach vehicles (ISO, 2022); IEEE 1547-2018 sets the interconnection and ride-through behaviour that any V2G discharge must respect (IEEE, 2018); IEEE C57.91 provides the transformer-loading guide against which the overload penalties of the testbed would be replaced by a thermal-ageing model in practice (IEEE, 2011); and the ITU-R framework for IMT-2030 defines the capabilities (hyper-reliable low-latency communication among them) that the 6G archetype stylises (ITU, 2023). The regulatory thread is reserve qualification: fast frequency products with latency and reliability requirements that only URLLC links meet are the mechanism by which connectivity acquires market value in this study (Han et al., 2010; Noor-A-Rahim et al., 2022), and whether such products are open to aggregated vehicle fleets is a decision for system operators and regulators. Governance of a learned controller needs a record: the data provenance (which telemetry, at what age), the safety constraints that bound the agent's actions, the seeds and held-out days on which it was evaluated, the conditions under which it was trained (the grid showed what happens outside them), and the measured outcomes that return to retraining (the governance lane of Figure 1 and Phase 4 of Figure 12) in line with the design-science and risk-management conventions the framework draws on (Gregor & Hevner, 2013; ISO, 2018; Rudin, 2019). Figure 12 sets out the implementation as four phases (data and connectivity, learn and validate, deploy with safety, scale and govern) with the governance and enablers each needs.

Implications for analytics practice

The study illustrates five analytics practices for learned controllers. Evaluate against a bound and a heuristic, with seeds and held-out days, and report distributions over days rather than point estimates: that every seed of the learned policy beat the heuristic under 6G and none under 4G carries more decision content than a mean (Henderson et al., 2018; Agarwal et al., 2021; Spiegelhalter, 2017). Model the communication layer explicitly, so that connectivity has a value that can be decomposed into channels and tested for degradation (Bennis et al., 2018; Kaul et al., 2012). Separate the learning

problem's reward from the decision's economics: the policies are trained on a fixed reward and then valued across 2,000 economic scenarios, which is what lets the Monte Carlo analysis state a probability that the connectivity premium pays without retraining the agents (Bertsimas & Kallus, 2020; Elmachtoub & Grigas, 2022). Attribute the variance of the decision quantity to its inputs with Sobol indices and summarise the scenario grid with an interpretable metamodel, so that the operator knows which uncertainties to buy down (the reserve price) and which conditions to seek (photovoltaics and volatility) (Sobol', 2001; Saltelli et al., 2010; Rudin, 2019). And keep the simulation honest with verification checks, common random numbers, a generalisation test outside the training distribution and a results file from which every table and figure is produced (Law, 2015; Sargent, 2013). These practices are the analytics counterpart of the engineering content; they are what turns a reinforcement-learning result into a decision.

Figure 12: Implementation roadmap for 6G-enabled V2G coordination: four phases from data and connectivity through learning and validation and deployment with safety to scaling and governance, with the governance bodies and enablers



Limitations

Table 15 lists the assumptions and their likely effects. The testbed is stylised throughout: no measured sessions, loads, prices, renewable profiles or network measurements were used, so absolute costs and savings describe the testbed only (the large savings against uncontrolled charging follow from a baseline that overloads a transformer and from a strong duck curve) and relative orderings (learned policies between the heuristic and the bound under 6G, the 6G effect split between market and information channels, savings rising with photovoltaics and volatility) are the findings to carry away, and even those depend on ranges that an operator should replace with its own data. The communication archetypes are representations of orders of magnitude from the URLLC and 6G literature, not measurements of any network, and the value of connectivity is as sensitive to the reserve-eligibility and reserve-price assumptions as the decomposition and the Sobol indices show. The learners are small NumPy networks trained for 500 episodes with three seeds on two cores; their validation curves had flattened but not converged to the bound, so the reported policies are

undertrained relative to what the literature's larger methods achieve, and the gap to the bound overstates what a tuned learner would leave. Hub-level actions with a heuristic allocation inside the hub lose vehicle-level optimisation, while the oracle relaxes to aggregated envelopes, so the bound is loose in the other direction. The learned policies did not generalise to a feeder without photovoltaics, and the grid did not retrain them. Degradation is a constant cost per kWh of throughput; reserve is paid on retained headroom without activation; the economics are in baseline-year terms without discounting or capital cost; and users are assumed to accept the departure targets and V2G floors as stated. The evaluation protocol (40 days, three seeds, 12 days per grid cell) leaves interval estimates wide, and differences of a few per cent between policies (VDN against the heuristic under 5G, for instance) are not resolved.

Table 15: Assumptions and limitations, their likely effect on the results and how to test them with real data

Assumption or limitation	Likely effect on results	How to test with real data
Stylised testbed: no measured loads, prices, sessions or renewables	Absolute costs and savings describe the testbed only; relative orderings are more robust than magnitudes	Replace generators with ACN-Data-style sessions, metered base loads, market prices and reanalysis PV/wind for the target feeder
Communication archetypes are representations, not measurements	The value of connectivity depends on the loss, delay and eligibility values assumed; the sweep bounds the sensitivity	Measure latency, loss and message age per hub; obtain reserve-qualification rules from the system operator
Small MARL at modest training budgets (NumPy MLP, 500 episodes, 3 seeds)	Learned policies may be far from the achievable optimum; the gap to the oracle overstates what a tuned learner would leave	Larger networks, longer training, QMIX/MAPPO, curriculum from the heuristic; report seeds and bounds
Hub-level actions with heuristic allocation inside the hub	Loses vehicle-level optimisation; the oracle relaxes to aggregated envelopes, so the bound is loose in the other direction	Vehicle-level agents or a hub-level LP inside the hub; tighter bound from a full MILP on small days
Linearised voltage and single-transformer feeder	Voltage violations are rare and penalties small; feeder losses are not modelled	Power-flow tool (pandapower, MATPOWER) on a test feeder; transformer thermal ageing (IEEE C57.91)
Degradation as a constant cost per kWh of throughput	Ignores SOC-window and temperature effects that can raise or lower V2G cost	Cycle-ageing model from the degradation literature calibrated to the fleet's chemistry
Reserve paid on retained headroom without activation	Reserve revenue is an upper bound on what activation would allow	Model activation events and their energy effects; market data on clearing prices
Economics in baseline-year terms without discounting or capital costs	Net value of 6G compares operating savings with a connectivity premium only	Life-cycle costing with charger, communication and platform capital; discount rate
Users assumed to accept V2G and departure targets as stated	User disutility, opt-outs and plug-in rates would reduce the available flexibility	Consent and incentive design; measured plug-in and opt-out rates
Evaluation protocol: 40 days, 3 seeds, 12 days per grid cell	Interval estimates are wide; small differences between policies are not resolved	More days and seeds; paired bootstrap; performance profiles

Further research

Five directions follow. First, calibration to a real feeder and fleet: ACN-Data-style sessions, metered base loads, market prices and reanalysis renewable profiles would replace the generators, and a power-flow tool on a test feeder would replace the linearised voltage model (Lee et al., 2019; Thurner et al., 2018). Second, stronger learners under the same evaluation protocol (QMIX, MAPPO, continuous actions, curricula from the heuristic, longer training, and training across the scenario grid so that policies generalise) to close the gap to the bound and to test whether the value of connectivity rises with the learner's capability (Rashid et al., 2018; Yu et al., 2022; Lowe et al., 2017). Third, a richer communication model: measured latency and loss distributions, queueing at the edge, federated or decentralised training over the wireless network, and integrated sensing, which the 6G literature anticipates (Chen et al., 2021; Park et al., 2019; Saad et al., 2020). Fourth, markets: activation of reserves with its energy effects, bidding under uncertainty, and the interaction of many aggregators on one feeder (Vagropoulos & Bakirtzis, 2013; Sortomme & El-Sharkawi, 2012). Fifth, people: plug-in behaviour, consent, incentives and the social dimensions of a V2G transition, which bound the flexibility that any controller can use (Sovacool et al., 2018; Noel et al., 2019). Each would use the same framework; only the layer would change.

Conclusion

This article asked what a decision-analytics framework for 6G-enabled multi-agent reinforcement learning in V2G coordination should contain, and what it says on a transparent testbed. The framework has four layers (physical, communication and edge, multi-agent learning, and decision, market and governance) and meets six requirements drawn from the V2G, charging-coordination, MARL, 6G and analytics literatures: constraints that bind, communication as a modelled layer with a value, learning with safety and coordination, evaluation against a bound with seeds and held-out days, decision outputs as probabilities and attributions, and feedback with a governance record. On a stylised feeder with eight hubs and 200 vehicles, every coordinated policy removed the transformer overload of uncontrolled charging; the learned policies beat a capacity-aware heuristic only when the network supplied fleet-flexibility messages and reserve eligibility, by 2.5% (VDN) and 7.8% (IQL) under 6G, while remaining 30–41% above a perfect-information bound; the value of 6G over 4G to the learned policy was 613 per vehicle-year gross and 552 net of the connectivity premium, positive in every economic draw, split between a market channel and an information channel of about 200 per day each, with the reserve price explaining 87% of its variance; and the saving rose with photovoltaics and price volatility, failed to generalise to a feeder without photovoltaics, and met a binding transformer at high penetration. The practical conclusion is that the case for 6G in V2G coordination rests on market access and on the edge aggregator's information as much as on the radio link, that learned controllers need a bound, a heuristic fallback and a generalisation test before deployment, and that the framework's outputs (probabilities, decompositions, Sobol indices and rules) are what an operator needs to decide. Every input is stylised and every number should be replaced with the operator's own; the testbed and results file are provided so that this can be done.

REFERENCES

[1]. Agarwal, R., Schwarzer, M., Castro, P. S., Courville, A., & Bellemare, M. G. (2021). Deep reinforcement learning at the edge of the statistical precipice. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)* (pp. 29304–29320). Curran Associates. https://papers.nips.cc/paper_files/paper/2021/hash/f514cec81cb148559cf475e7426eed5e-Abstract.html

[2]. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)* (pp. 1–13). ACM. https://doi.org/10.1145/3290605.3300233

[3]. Aminul Islam, M. (2024). Interpreting Cyber-Threat Indicators in Enterprise Memory Forensics: A Mixed-Methods Study of Machine Learning Performance and Explainability. *Review of Applied Science and Technology*, *14*(12), 01–13. https://doi.org/10.63125/2p8hte92

[4]. Aminul Islam, M., & Md, A. (2023). Adaptive Class-Imbalance Learning for High-Sensitivity Fraud Detection in Large-Scale Digital Transaction Networks. *International Journal of Business and Economics Insights*, *13*(09), 01-33. https://doi.org/10.63125/bmxwv741

[5]. Ban, G.-Y., & Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, *67*(1), 90–108. https://doi.org/10.1287/opre.2018.1757

[6]. Baran, M. E., & Wu, F. F. (1989). Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Transactions on Power Delivery*, *4*(2), 1401–1407. https://doi.org/10.1109/61.25627

[7]. Bennis, M., Debbah, M., & Poor, H. V. (2018). Ultrareliable and low-latency wireless communication: Tail, risk, and scale. *Proceedings of the IEEE*, *106*(10), 1834–1853. https://doi.org/10.1109/JPROC.2018.2867029

[8]. Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, *66*(3), 1025–1044. https://doi.org/10.1287/mnsc.2018.3253

[9]. Borenstein, S. (2005). The long-run efficiency of real-time electricity pricing. *The Energy Journal*, *26*(3), 93–116. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol26-No3-5

[10]. Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, *13*(1), 333–372. https://doi.org/10.1257/mac.20180386

[11]. Buonocore, J. J., Salimifard, P., Magavi, Z., & Allen, J. G. (2022). Inefficient building electrification will require massive buildout of renewable energy and seasonal energy storage. *Scientific Reports*, *12*, Article 11931. https://doi.org/10.1038/s41598-022-15628-2

[12]. Buşoniu, L., Babuška, R., & De Schutter, B. (2008). A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *38*(2), 156–172. https://doi.org/10.1109/TSMCC.2007.913919

[13]. Callaway, D. S., & Hiskens, I. A. (2011). Achieving controllability of electric loads. *Proceedings of the IEEE*, *99*(1), 184–199. https://doi.org/10.1109/JPROC.2010.2081652

[14]. Cao, D., Hu, W., Zhao, J., Zhang, G., Zhang, B., Liu, Z., Chen, Z., & Blaabjerg, F. (2020). Reinforcement learning and its applications in modern power and energy systems: A review. *Journal of Modern Power Systems and Clean Energy*, *8*(6), 1029–1042. https://doi.org/10.35833/MPCE.2020.000552

[15]. Castañeda Garcia, M. H., Molina-Galan, A., Boban, M., Gozalvez, J., Coll-Perales, B., Şahin, T., & Kousaridas, A. (2021). A tutorial on 5G NR V2X communications. *IEEE Communications Surveys & Tutorials*, *23*(3), 1972–2026. https://doi.org/10.1109/COMST.2021.3057017

[16]. Chen, M., Gündüz, D., Huang, K., Saad, W., Bennis, M., Feljan, A. V., & Poor, H. V. (2021). Distributed learning in wireless networks: Recent progress and future challenges. *IEEE Journal on Selected Areas in Communications*, *39*(12), 3579–3605. https://doi.org/10.1109/JSAC.2021.3118346

[17]. Chen, X., Qu, G., Tang, Y., Low, S. H., & Li, N. (2022). Reinforcement learning for selective key applications in power systems: Recent advances and future challenges. *IEEE Transactions on Smart Grid*, *13*(4), 2935–2958. https://doi.org/10.1109/TSG.2022.3154718

[18]. Clement-Nyns, K., Haesen, E., & Driesen, J. (2010). The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. *IEEE Transactions on Power Systems*, *25*(1), 371–380. https://doi.org/10.1109/TPWRS.2009.2036481

[19]. da Silva, F. L., Nishida, C. E. H., Roijers, D. M., & Costa, A. H. R. (2020). Coordination of electric vehicle charging through multiagent reinforcement learning. *IEEE Transactions on Smart Grid*, *11*(3), 2347–2356. https://doi.org/10.1109/TSG.2019.2952331

[20]. Dang, S., Amin, O., Shihada, B., & Alouini, M.-S. (2020). What should 6G be? *Nature Electronics*, *3*(1), 20–29. https://doi.org/10.1038/s41928-019-0355-6

[21]. Deng, R., Yang, Z., Chow, M.-Y., & Chen, J. (2015). A survey on demand response in smart grids: Mathematical models and approaches. *IEEE Transactions on Industrial Informatics*, *11*(3), 570–582. https://doi.org/10.1109/TII.2015.2414719

[22]. Denholm, P., & Hand, M. (2011). Grid flexibility and storage required to achieve very high penetration of variable renewable electricity. *Energy Policy*, *39*(3), 1817–1830. https://doi.org/10.1016/j.enpol.2011.01.019

[23]. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning (arXiv:1702.08608). arXiv. https://arxiv.org/abs/1702.08608

[24]. Dubarry, M., Devie, A., & McKenzie, K. (2017). Durability and reliability of electric vehicle batteries under electric utility grid operations: Bidirectional charging impact analysis. *Journal of Power Sources*, *358*, 39–49. https://doi.org/10.1016/j.jpowsour.2017.05.015

[25]. Dulac-Arnold, G., Levine, N., Mankowitz, D. J., Li, J., Paduraru, C., Gowal, S., & Hester, T. (2021). Challenges of real-world reinforcement learning: Definitions, benchmarks and analysis. *Machine Learning*, *110*(9), 2419–2468. https://doi.org/10.1007/s10994-021-05961-4

[26]. Elmachtoub, A. N., & Grigas, P. (2022). Smart "predict, then optimize". *Management Science*, *68*(1), 9–26. https://doi.org/10.1287/mnsc.2020.3922

[27]. Fang, X., Misra, S., Xue, G., & Yang, D. (2012). Smart grid, the new and improved power grid: A survey. *IEEE Communications Surveys & Tutorials*, *14*(4), 944–980. https://doi.org/10.1109/SURV.2011.101911.00087

[28]. Farivar, M., & Low, S. H. (2013). Branch flow model: Relaxations and convexification, Part I. *IEEE Transactions on Power Systems*, *28*(3), 2554–2564. https://doi.org/10.1109/TPWRS.2013.2255317

[29]. Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, *25*(1), 3–23. https://doi.org/10.1016/j.ijforecast.2008.11.010

[30]. Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, *20*(177), 1–81. https://jmlr.org/papers/v20/18-760.html

[31]. Foerster, J., Farquhar, G., Afouras, T., Nardelli, N., & Whiteson, S. (2018). Counterfactual multi-agent policy gradients. In *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 2974–2982). AAAI Press. https://doi.org/10.1609/aaai.v32i1.11794

[32]. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

[33]. Gagnon, P., Frazier, W., Cole, W., & Hale, E. (2021). *Cambium documentation: Version 2021* (NREL/TP-6A40-81611). National Renewable Energy Laboratory. https://www.nrel.gov/docs/fy22osti/81611.pdf

[34]. Gan, L., Topcu, U., & Low, S. H. (2013). Optimal decentralized protocol for electric vehicle charging. *IEEE Transactions on Power Systems*, *28*(2), 940–951. https://doi.org/10.1109/TPWRS.2012.2210288

[35]. Giordani, M., Polese, M., Mezzavilla, M., Rangan, S., & Zorzi, M. (2020). Toward 6G networks: Use cases and technologies. *IEEE Communications Magazine*, *58*(3), 55–61. https://doi.org/10.1109/MCOM.001.1900411

[36]. Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

[37]. Guille, C., & Gross, G. (2009). A conceptual framework for the vehicle-to-grid (V2G) implementation. *Energy Policy*, *37*(11), 4379–4390. https://doi.org/10.1016/j.enpol.2009.05.053

[38]. Gungor, V. C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., & Hancke, G. P. (2011). Smart grid technologies: Communication technologies and standards. *IEEE Transactions on Industrial Informatics*, *7*(4), 529–539. https://doi.org/10.1109/TII.2011.2166794

[39]. Han, S., Han, S., & Sezaki, K. (2010). Development of an optimal vehicle-to-grid aggregator for frequency regulation. *IEEE Transactions on Smart Grid*, *1*(1), 65–72. https://doi.org/10.1109/TSG.2010.2045163

[40]. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

[41]. Haug, A., Zachariassen, F., & van Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management*, *4*(2), 168–193. https://doi.org/10.3926/jiem.2011.v4n2.p168-193

[42]. Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*, 72–80. https://doi.org/10.1016/j.ijpe.2014.04.018

[43]. Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., & Meger, D. (2018). Deep reinforcement learning that matters. In *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 3207–3214). AAAI Press. https://doi.org/10.1609/aaai.v32i1.11694

[44]. Herman, J., & Usher, W. (2017). SALib: An open-source Python library for sensitivity analysis. *Journal of Open Source Software*, *2*(9), Article 97. https://doi.org/10.21105/joss.00097

[45]. Hernández-Leal, P., Kartal, B., & Taylor, M. E. (2019). A survey and critique of multiagent deep reinforcement learning. *Autonomous Agents and Multi-Agent Systems*, *33*(6), 750–797. https://doi.org/10.1007/s10458-019-09421-1

[46]. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

[47]. Hilshey, A. D., Hines, P. D. H., Rezaei, P., & Dowds, J. R. (2013). Estimating the impact of electric vehicle smart charging on distribution transformer aging. *IEEE Transactions on Smart Grid*, *4*(2), 905–913. https://doi.org/10.1109/TSG.2012.2217385

[48]. Hirth, L. (2013). The market value of variable renewables. *Energy Economics*, *38*, 218–236. https://doi.org/10.1016/j.eneco.2013.02.004

[49]. Howard, R. A. (1966). Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, *2*(1), 22–26. https://doi.org/10.1109/TSSC.1966.300074

[50]. Hu, J., Morais, H., Sousa, T., & Lind, M. (2016). Electric vehicle fleet management in smart grids: A review of services, optimization and control aspects. *Renewable and Sustainable Energy Reviews*, *56*, 1207–1226. https://doi.org/10.1016/j.rser.2015.12.014

[51]. Hui, H., Ding, Y., Shi, Q., Li, F., Song, Y., & Yan, J. (2020). 5G network-based Internet of Things for demand response in smart grid: A survey on application potential. *Applied Energy*, *257*, Article 113972. https://doi.org/10.1016/j.apenergy.2019.113972

[52]. Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, *110*(3), 457–506. https://doi.org/10.1007/s10994-021-05946-3

[53]. IEEE. (2011). *IEEE Std C57.91-2011: IEEE guide for loading mineral-oil-immersed transformers and step-voltage regulators*. IEEE. https://doi.org/10.1109/IEEESTD.2012.6166928

[54]. IEEE. (2018). *IEEE Std 1547-2018: IEEE standard for interconnection and interoperability of distributed energy resources with associated electric power systems interfaces*. IEEE. https://doi.org/10.1109/IEEESTD.2018.8332112

[55]. International Energy Agency. (2024). *Global EV Outlook 2024: Moving towards increased affordability*. IEA. https://www.iea.org/reports/global-ev-outlook-2024

[56]. International Organization for Standardization. (2018). *ISO 31000:2018 Risk management: Guidelines*. ISO. https://www.iso.org/standard/65694.html

[57]. International Organization for Standardization. (2022). *ISO 15118-20:2022 Road vehicles, vehicle to grid communication interface, part 20: 2nd generation network layer and application layer requirements*. ISO. https://www.iso.org/standard/77845.html

[58]. International Telecommunication Union. (2023). *Recommendation ITU-R M.2160-0: Framework and overall objectives of the future development of IMT for 2030 and beyond*. ITU. https://www.itu.int/rec/R-REC-M.2160-0-202311-I/en

[59]. Iwanaga, T., Usher, W., & Herman, J. (2022). Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. *Socio-Environmental Systems Modelling*, *4*, Article 18155. https://doi.org/10.18174/sesmo.18155

[60]. Jenkins, J. D., Luke, M., & Thernstrom, S. (2018). Getting to zero carbon emissions in the electric power sector. *Joule*, *2*(12), 2498–2510. https://doi.org/10.1016/j.joule.2018.11.013

[61]. Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux. https://openlibrary.org/isbn/9780374275631

[62]. Kalalas, C., Thrybom, L., & Alonso-Zarate, J. (2016). Cellular communications for smart grid neighborhood area networks: A survey. *IEEE Access*, *4*, 1469–1493. https://doi.org/10.1109/ACCESS.2016.2551978

[63]. Kaul, S. K., Yates, R. D., & Gruteser, M. (2012). Real-time status: How often should one update? In *2012 Proceedings IEEE INFOCOM* (pp. 2731–2735). IEEE. https://doi.org/10.1109/INFCOM.2012.6195689

[64]. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*, 3146–3154. https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree

[65]. Kempton, W., & Tomić, J. (2005a). Vehicle-to-grid power fundamentals: Calculating capacity and net revenue. *Journal of Power Sources*, *144*(1), 268–279. https://doi.org/10.1016/j.jpowsour.2004.12.025

[66]. Kempton, W., & Tomić, J. (2005b). Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy. *Journal of Power Sources*, *144*(1), 280–294. https://doi.org/10.1016/j.jpowsour.2004.12.022

[67]. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*. Curran Associates. https://proceedings.neurips.cc/paper_files/paper/2017/hash/2650d6089a6d640c5e85b2b88265dc2b-Abstract.html

[68]. Kersting, W. H. (1991). Radial distribution test feeders. *IEEE Transactions on Power Systems*, *6*(3), 975–985. https://doi.org/10.1109/59.119237

[69]. Kynatun, M., & Saba, A. (2024). Machine Learning Enhanced Intrusion Detection and Anomaly Classification for MIS-Driven Enterprise Network Security. *International Journal of Scientific Interdisciplinary Research*, *14*(09), 01–41. https://doi.org/10.63125/xk6kpc72

[70]. Law, A. M. (2015). *Simulation modeling and analysis* (5th ed.). McGraw-Hill Education. https://www.mheducation.com/

[71]. Lee, Z. J., Li, T., & Low, S. H. (2019). ACN-Data: Analysis and applications of an open EV charging dataset. In *Proceedings of the Tenth ACM International Conference on Future Energy Systems (e-Energy 2019)* (pp. 139–149). ACM. https://doi.org/10.1145/3307772.3328313

[72]. Lee, Z. J., Sharma, S., Johansson, D., & Low, S. H. (2021). ACN-Sim: An open-source simulator for data-driven electric vehicle charging research. *IEEE Transactions on Smart Grid*, *12*(6), 5113–5123. https://doi.org/10.1109/TSG.2021.3103156

[73]. Letaief, K. B., Chen, W., Shi, Y., Zhang, J., & Zhang, Y.-J. A. (2019). The roadmap to 6G: AI empowered wireless networks. *IEEE Communications Magazine*, *57*(8), 84–90. https://doi.org/10.1109/MCOM.2019.1900271

[74]. Li, H., Wan, Z., & He, H. (2020). Constrained EV charging scheduling based on safe deep reinforcement learning. *IEEE Transactions on Smart Grid*, *11*(3), 2427–2439. https://doi.org/10.1109/TSG.2019.2955437

[75]. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). *Continuous control with deep reinforcement learning* (arXiv:1509.02971). arXiv. https://arxiv.org/abs/1509.02971

[76]. Lipton, Z. C. (2018). The mythos of model interpretability. *Queue, 16*(3), 31–57. https://doi.org/10.1145/3236386.3241340

[77]. Lopes, J. A. P., Soares, F. J., & Almeida, P. M. R. (2011). Integration of electric vehicles in the electric power system. *Proceedings of the IEEE*, *99*(1), 168–183. https://doi.org/10.1109/JPROC.2010.2066250

[78]. Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*. Curran Associates. https://papers.nips.cc/paper/7217-multi-agent-actor-critic-for-mixed-cooperative-competitive-environments

[79]. Lund, P. D., Lindgren, J., Mikkola, J., & Salpakari, J. (2015). Review of energy system flexibility measures to enable high levels of variable renewable electricity. *Renewable and Sustainable Energy Reviews*, *45*, 785–807. https://doi.org/10.1016/j.rser.2015.01.057

[80]. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774. https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

[81]. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

[82]. Ma, Z., Callaway, D. S., & Hiskens, I. A. (2013). Decentralized charging control of large populations of plug-in electric vehicles. *IEEE Transactions on Control Systems Technology*, *21*(1), 67–78. https://doi.org/10.1109/TCST.2011.2174059

[83]. Mai, T., Jadun, P., Logan, J., McMillan, C., Muratori, M., Steinberg, D., Vimmerstedt, L., Jones, R., Haley, B., & Nelson, B. (2018). *Electrification futures study: Scenarios of electric technology adoption and power consumption for the United States* (NREL/TP-6A20-71500). National Renewable Energy Laboratory. https://doi.org/10.2172/1459351

[84]. Mathieu, J. L., Koch, S., & Callaway, D. S. (2013). State estimation and control of electric loads to manage real-time energy imbalance. *IEEE Transactions on Power Systems*, *28*(1), 430–440. https://doi.org/10.1109/TPWRS.2012.2204074

[85]. Mansura Akter, E., & Sumaiya, M. (2022). Antimicrobial Resistance Profiles of Acinetobacter Baumannii and Vibrio Cholerae in Hospital Wastewater and Adjacent Household Water Sources. *American Journal of Data Science and Analytics*, *3*(09), 01-55. https://doi.org/10.63125/gbervt14

[86]. Md. Morshedul, I., & Aminul Islam, M. (2022). A Quantitative Evaluation of Explainable Ensemble Learning for Probabilistic Risk Estimation and Decision Reliability in Enterprise Analytics. *American Journal of Data Science and Analytics*, *3*(02), 01-38. https://doi.org/10.63125/k22shk74

[87]. Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, *98*, 261–276. https://doi.org/10.1016/j.jbusres.2019.01.044

[88]. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236

[89]. Molnar, C. (2025). *Interpretable machine learning: A guide for making black box models explainable* (3rd ed.). Independently published. https://christophm.github.io/interpretable-ml-book/

[90]. Morgan, M. G., & Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511840609

[91]. Mukherjee, J. C., & Gupta, A. (2015). A review of charge scheduling of electric vehicles in smart grid. *IEEE Systems Journal*, *9*(4), 1541–1553. https://doi.org/10.1109/JSYST.2014.2356559

[92]. Mwasilu, F., Justo, J. J., Kim, E.-K., Do, T. D., & Jung, J.-W. (2014). Electric vehicles and smart grid interaction: A review on vehicle to grid and renewable energy sources integration. *Renewable and Sustainable Energy Reviews*, *34*, 501–516. https://doi.org/10.1016/j.rser.2014.03.031

[93]. Noel, L., Zarazua de Rubens, G., Kester, J., & Sovacool, B. K. (2019). *Vehicle-to-grid: A sociotechnical transition beyond electric mobility*. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-04864-8

[94]. Noor-A-Rahim, M., Liu, Z., Lee, H., Khyam, M. O., He, J., Pesch, D., Moessner, K., Saad, W., & Poor, H. V. (2022). 6G for vehicle-to-everything (V2X) communications: Enabling technologies, challenges, and opportunities. *Proceedings of the IEEE*, *110*(6), 712–734. https://doi.org/10.1109/JPROC.2022.3173031

[95]. O'Connell, N., Pinson, P., Madsen, H., & O'Malley, M. J. (2014). Benefits and challenges of electrical demand response: A critical review. *Renewable and Sustainable Energy Reviews*, *39*, 686–699. https://doi.org/10.1016/j.rser.2014.07.098

[96]. Oroojlooy, A., & Hajinezhad, D. (2023). A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence*, *53*(11), 13677–13722. https://doi.org/10.1007/s10489-022-04105-y

[97]. Palensky, P., & Dietrich, D. (2011). Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE Transactions on Industrial Informatics*, *7*(3), 381–388. https://doi.org/10.1109/TII.2011.2158841

[98]. Park, J., Samarakoon, S., Bennis, M., & Debbah, M. (2019). Wireless network intelligence at the edge. *Proceedings of the IEEE*, *107*(11), 2204–2239. https://doi.org/10.1109/JPROC.2019.2941458

[99]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

[100]. Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

[101]. Perera, A. T. D., & Kamalaruban, P. (2021). Applications of reinforcement learning in energy systems. *Renewable and Sustainable Energy Reviews*, *137*, Article 110618. https://doi.org/10.1016/j.rser.2020.110618

[102]. Peterson, S. B., Whitacre, J. F., & Apt, J. (2010). The economics of using plug-in hybrid electric vehicle battery packs for grid storage. *Journal of Power Sources*, *195*(8), 2377–2384. https://doi.org/10.1016/j.jpowsour.2009.09.070

[103]. Pfenninger, S., & Staffell, I. (2016). Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. *Energy*, *114*, 1251–1265. https://doi.org/10.1016/j.energy.2016.08.060

[104]. Popovski, P., Stefanović, Č., Nielsen, J. J., de Carvalho, E., Angjelichinoski, M., Trillingsgaard, K. F., & Bana, A.-S. (2019). Wireless access in ultra-reliable low-latency communication (URLLC). *IEEE Transactions on Communications*, *67*(8), 5783–5801. https://doi.org/10.1109/TCOMM.2019.2914652

[105]. Rashid, T., Samvelyan, M., Schroeder de Witt, C., Farquhar, G., Foerster, J., & Whiteson, S. (2018). QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. In *Proceedings of the 35th International Conference on Machine Learning* (pp. 4295–4304). PMLR. https://proceedings.mlr.press/v80/rashid18a.html

[106]. Reka, S. S., Dragičević, T., Siano, P., & Prabaharan, S. R. S. (2019). Future generation 5G wireless networks for smart grid: A comprehensive review. *Energies*, *12*(11), Article 2140. https://doi.org/10.3390/en12112140

[107]. Rennert, K., Errickson, F., Prest, B. C., Rennels, L., Newell, R. G., Pizer, W., Kingdon, C., Wingenroth, J., Cooke, R., Parthum, B., Smith, D., Cromar, K., Diaz, D., Moore, F. C., Müller, U. K., Plevin, R. J., Raftery, A. E., Ševčíková, H., Sheets, H., ... Anthoff, D. (2022). Comprehensive evidence implies a higher social cost of CO2. *Nature*, *610*(7933), 687–692. https://doi.org/10.1038/s41586-022-05224-9

[108]. Richardson, D. B. (2013). Electric vehicles and the electric grid: A review of modeling approaches, impacts, and renewable energy integration. *Renewable and Sustainable Energy Reviews*, *19*, 247–254. https://doi.org/10.1016/j.rser.2012.11.042

[109]. Rolf, B., Jackson, I., Müller, M., Lang, S., Reggelin, T., & Ivanov, D. (2023). A review on reinforcement learning algorithms and applications in supply chain management. *International Journal of Production Research*, *61*(20), 7151–7179. https://doi.org/10.1080/00207543.2022.2140221

[110]. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[111]. Saad, W., Bennis, M., & Chen, M. (2020). A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. *IEEE Network*, *34*(3), 134–142. https://doi.org/10.1109/MNET.001.1900287

[112]. Sadeghianpourhamami, N., Deleu, J., & Develder, C. (2020). Definition and evaluation of model-free coordination of electrical vehicle charging with reinforcement learning. *IEEE Transactions on Smart Grid*, *11*(1), 203–214. https://doi.org/10.1109/TSG.2019.2920320

[113]. Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, *181*(2), 259–270. https://doi.org/10.1016/j.cpc.2009.09.018

[114]. Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation*, *7*(1), 12–24. https://doi.org/10.1057/jos.2012.20

[115]. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal policy optimization algorithms* (arXiv:1707.06347). arXiv. https://arxiv.org/abs/1707.06347

[116]. Shin, M., Choi, D.-H., & Kim, J. (2020). Cooperative management for PV/ESS-enabled electric vehicle charging stations: A multiagent deep reinforcement learning approach. *IEEE Transactions on Industrial Informatics*, *16*(5), 3493–3503. https://doi.org/10.1109/TII.2019.2944183

[117]. Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

[118]. Siano, P. (2014). Demand response and smart grids: A survey. *Renewable and Sustainable Energy Reviews*, *30*, 461–478. https://doi.org/10.1016/j.rser.2013.10.022

[119]. Siler-Evans, K., Azevedo, I. L., & Morgan, M. G. (2012). Marginal emissions factors for the U.S. electricity system. *Environmental Science & Technology*, *46*(9), 4742–4748. https://doi.org/10.1021/es300145v

[120]. Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, *55*(1-3), 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6

[121]. Sonnenberg, C., & vom Brocke, J. (2012). Evaluations in the science of the artificial – Reconsidering the build-evaluate pattern in design science research. In *Design Science Research in Information Systems. Advances in Theory and Practice (DESRIST 2012), Lecture Notes in Computer Science, Vol. 7286* (pp. 381–397). Springer. https://doi.org/10.1007/978-3-642-29863-9_28

[122]. Sortomme, E., & El-Sharkawi, M. A. (2011). Optimal charging strategies for unidirectional vehicle-to-grid. *IEEE Transactions on Smart Grid*, *2*(1), 131–138. https://doi.org/10.1109/TSG.2010.2090910

[123]. Sortomme, E., & El-Sharkawi, M. A. (2012). Optimal scheduling of vehicle-to-grid energy and ancillary services. *IEEE Transactions on Smart Grid*, *3*(1), 351–359. https://doi.org/10.1109/TSG.2011.2164099

[124]. Sovacool, B. K., Noel, L., Axsen, J., & Kempton, W. (2018). The neglected social dimensions to a vehicle-to-grid (V2G) transition: A critical and systematic review. *Environmental Research Letters*, *13*(1), Article 013001. https://doi.org/10.1088/1748-9326/aa9c6d

[125]. Spiegelhalter, D. (2017). Risk and uncertainty communication. *Annual Review of Statistics and Its Application*, *4*, 31–60. https://doi.org/10.1146/annurev-statistics-010814-020148

[126]. Staffell, I., & Pfenninger, S. (2016). Using bias-corrected reanalysis to simulate current and future wind power output. *Energy*, *114*, 1224–1239. https://doi.org/10.1016/j.energy.2016.08.068

[127]. Sunehag, P., Lever, G., Gruslys, A., Czarnecki, W. M., Zambaldi, V., Jaderberg, M., Lanctot, M., Sonnerat, N., Leibo, J. Z., Tuyls, K., & Graepel, T. (2018). Value-decomposition networks for cooperative multi-agent learning based on team reward. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2018)* (pp. 2085–2087). International Foundation for Autonomous Agents and Multiagent Systems. https://www.ifaamas.org/Proceedings/aamas2018/pdfs/p2085.pdf

[128]. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press. http://incompleteideas.net/book/the-book-2nd.html

[129]. Sumaiya, M. (2024). Hospital Wastewater as A Reservoir of Antimicrobial Resistance Genes: Implications for Public Health and Genomic Surveillance in The United States. Review of Applied Science and Technology, 3(04), 448–508. https://doi.org/10.63125/ahvas976

[130]. Sumaiya, M. (2025). CRISPR-Cas Systems in Precision Oncology: Current Applications and Future Challenges. American Journal of Interdisciplinary Studies, 6(1), 513-564. https://doi.org/10.63125/rqpqr418

[131]. Sumaiya, M. (2026). Artificial Intelligence Driven Multi-Omics Approaches for Precision Medicine: From Molecular Discovery to Clinical Translation. American Journal of Scholarly Research and Innovation, 5(01), 429–477. https://doi.org/10.63125/yfme0q12

[132]. Tan, K. M., Ramachandaramurthy, V. K., & Yong, J. Y. (2016). Integration of electric vehicles in smart grid: A review on vehicle to grid technologies and optimization techniques. *Renewable and Sustainable Energy Reviews*, *53*, 720–732. https://doi.org/10.1016/j.rser.2015.09.012

[133]. Tan, M. (1993). Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Machine Learning Proceedings 1993 (Proceedings of the Tenth International Conference on Machine Learning)* (pp. 330–337). Morgan Kaufmann. https://doi.org/10.1016/B978-1-55860-307-3.50049-6

[134]. Tataria, H., Shafi, M., Molisch, A. F., Dohler, M., Sjöland, H., & Tufvesson, F. (2021). 6G wireless systems: Vision, requirements, challenges, insights, and opportunities. *Proceedings of the IEEE*, *109*(7), 1166–1199. https://doi.org/10.1109/JPROC.2021.3061701

[135]. Thurner, L., Scheidler, A., Schäfer, F., Menke, J.-H., Dollichon, J., Meier, F., Meinecke, S., & Braun, M. (2018). pandapower: An open-source Python tool for convenient modeling, analysis, and optimization of electric power systems. *IEEE Transactions on Power Systems*, *33*(6), 6510–6521. https://doi.org/10.1109/TPWRS.2018.2829021

[136]. Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

[137]. U.S. Environmental Protection Agency. (2023). *Report on the social cost of greenhouse gases: Estimates incorporating recent scientific advances*. U.S. Environmental Protection Agency. https://www.epa.gov/system/files/documents/2023-12/epa_scghg_2023_report_final.pdf

[138]. U.S. Environmental Protection Agency. (2024). *Emissions & Generation Resource Integrated Database (eGRID), 2022 data*. U.S. Environmental Protection Agency. https://www.epa.gov/egrid

[139]. Uddin, K., Jackson, T., Widanage, W. D., Chouchelamane, G., Jennings, P. A., & Marco, J. (2017). On the possibility of extending the lifetime of lithium-ion batteries through optimal V2G facilitated by an integrated vehicle and smart-grid system. *Energy*, *133*, 710–722. https://doi.org/10.1016/j.energy.2017.04.116

[140]. Vagropoulos, S. I., & Bakirtzis, A. G. (2013). Optimal bidding strategy for electric vehicle aggregators in electricity markets. *IEEE Transactions on Power Systems*, *28*(4), 4031–4041. https://doi.org/10.1109/TPWRS.2013.2274673

[141]. Vandael, S., Claessens, B., Ernst, D., Holvoet, T., & Deconinck, G. (2015). Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market. *IEEE Transactions on Smart Grid*, *6*(4), 1795–1805. https://doi.org/10.1109/TSG.2015.2393059

[142]. Vázquez-Canteli, J. R., & Nagy, Z. (2019). Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Applied Energy*, *235*, 1072–1089. https://doi.org/10.1016/j.apenergy.2018.11.002

[143]. Vázquez-Canteli, J. R., Kämpf, J., Henze, G., & Nagy, Z. (2019). CityLearn v1.0: An OpenAI Gym environment for demand response with deep reinforcement learning. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys 2019)* (pp. 356–357). ACM. https://doi.org/10.1145/3360322.3360998

[144]. Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems*, *25*(1), 77–89. https://doi.org/10.1057/ejis.2014.36

[145]. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

[146]. Waite, M., & Modi, V. (2020). Electricity load implications of space heating decarbonization pathways. *Joule*, *4*(2), 376–394. https://doi.org/10.1016/j.joule.2019.11.011

[147]. Wan, Z., Li, H., He, H., & Prokhorov, D. (2019). Model-free real-time EV charging scheduling based on deep reinforcement learning. *IEEE Transactions on Smart Grid*, *10*(5), 5246–5257. https://doi.org/10.1109/TSG.2018.2879572

[148]. Xu, B., Oudalov, A., Ulbig, A., Andersson, G., & Kirschen, D. S. (2018). Modeling of lithium-ion battery degradation for cell life assessment. *IEEE Transactions on Smart Grid*, *9*(2), 1131–1140. https://doi.org/10.1109/TSG.2016.2578950

[149]. Yan, Y., Qian, Y., Sharif, H., & Tipper, D. (2013). A survey on smart grid communication infrastructures: Motivations, requirements and challenges. *IEEE Communications Surveys & Tutorials*, *15*(1), 5–20. https://doi.org/10.1109/SURV.2012.021312.00034

[150]. Yang, T., Zhao, L., Li, W., & Zomaya, A. Y. (2020). Reinforcement learning in sustainable energy and electric systems: A survey. *Annual Reviews in Control*, *49*, 145–163. https://doi.org/10.1016/j.arcontrol.2020.03.001

[151]. Yates, R. D., Sun, Y., Brown, D. R., Kaul, S. K., Modiano, E., & Ulukus, S. (2021). Age of information: An introduction and survey. *IEEE Journal on Selected Areas in Communications*, *39*(5), 1183–1210. https://doi.org/10.1109/JSAC.2021.3065072

[152]. Yilmaz, M., & Krein, P. T. (2013). Review of the impact of vehicle-to-grid technologies on distribution systems and utility interfaces. *IEEE Transactions on Power Electronics*, *28*(12), 5673–5689. https://doi.org/10.1109/TPEL.2012.2227500

[153]. Yu, C., Velu, A., Vinitsky, E., Gao, J., Wang, Y., Bayen, A., & Wu, Y. (2022). The surprising effectiveness of PPO in cooperative multi-agent games. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)* (pp. 24611–24624). Curran Associates. https://doi.org/10.52202/068431-1787

[154]. Zhang, K., Yang, Z., & Başar, T. (2021). Multi-agent reinforcement learning: A selective overview of theories and algorithms. In *Handbook of reinforcement learning and control (Studies in Systems, Decision and Control, Vol. 325)* (pp. 321–384). Springer. https://doi.org/10.1007/978-3-030-60990-0_12

[155]. Zhang, Z., Xiao, Y., Ma, Z., Xiao, M., Ding, Z., Lei, X., Karagiannidis, G. K., & Fan, P. (2019). 6G wireless networks: Vision, requirements, architecture, and key technologies. *IEEE Vehicular Technology Magazine*, *14*(3), 28–41. https://doi.org/10.1109/MVT.2019.2921208

[156]. Zhang, Z., Zhang, D., & Qiu, R. C. (2020). Deep reinforcement learning for power system applications: An overview. *CSEE Journal of Power and Energy Systems*, *6*(1), 213–225. https://doi.org/10.17775/CSEEJPES.2019.00920

[157]. Zimmerman, R. D., Murillo-Sánchez, C. E., & Thomas, R. J. (2011). MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, *26*(1), 12–19. https://doi.org/10.1109/TPWRS.2010.2051168