Proceedings of CONF-MLA 2026 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2026.LD37913

Machine Learning for Virtual Power Plant Scheduling: From Deliverable Flexibility to Network-Constrained Operation

Zhengxuan Zhong

James Watt School of Engineering, University of Glasgow, Glasgow, UK
zhongzhengxuan25@outlook.com

Abstract. The capacity available to a virtual power plant can change between a market offer and delivery. This review examines fifteen studies published between 2019 and 2026, asking how that capacity is estimated and used in bidding, dispatch, and network coordination. For building demand, a reduction in consumption changes the thermal state and may require recovery later. The amount assigned now therefore affects what remains available for later requests. Learning can represent such dependencies through forecasts and resource models supplied to an optimizer, or through policies that select actions directly. In the bidding comparison, the larger forecast-based profit is also the one further from the calculation using actual data; it cannot be read as a higher realized return. Network coordination can reduce operator costs while increasing those of a participating virtual power plant, and differences in power-flow models complicate the comparison. What remains unresolved in the evidence assembled here is whether estimated flexibility stays deliverable as resources change. A record of the estimate, device assignments, and measured response for the same request would help locate a shortfall. It could distinguish a capacity estimate that has become outdated from a feasible allocation that was not carried out, directing attention to the part of the process that needs correction.

Keywords: virtual power plant, flexibility aggregation, electricity-market bidding, reinforcement learning, distribution-network coordination

1. Introduction

A virtual power plant (VPP) offers a service as a single market participant using equipment that remains physically dispersed. The capacity it can deliver at a particular time depends on the condition and availability of those resources. In the Northern European implementation described by Subramanya et al., measurements and forecasts obtained through interfaces to a commercial VPP support participation in primary frequency reserves [1].

Resource availability is also affected by earlier decisions. Reducing a building's demand changes its thermal state and can create a recovery requirement later in the schedule [2]. Customer participation adds an economic consideration: an incentive payment can alter the response available to the operator [3]. An offer based only on current availability can therefore miss both the later recovery of demand and changes in customer participation.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.LD37913

The fifteen studies considered here connect learning to successive decisions in VPP operation. In work on building dynamics and customer incentives, estimating capacity involves both the physical response and the willingness to supply it; cluster prediction adds an account of expected availability [2-4]. Bidding studies examine how forecasts affect a market decision, while decision-focused training allows dispatch costs to influence the predictor [5, 6]. Direct allocation policies cover regulation, residential schedules, and deadline-based tasks [7-10]. Their decisions also have a network setting: several VPPs may interact through prices or share constraints, prompting multi-agent and constraint-aware designs [11-15]. Reading these studies along the path to delivery exposes the capacity assumptions passed from one decision to the next.

After the review method in Section 2, Section 3 examines the 2019-2026 studies by operating task. This provides the basis for the comparisons in Section 4 and the conclusions and future questions in Section 5. The analysis addresses four questions:

RQ1: Which decisions between resource assessment and VPP service delivery use learning?

RQ2: What do the reported experiments show about economic performance, technical outcomes, and computational demand?

RQ3: How do resource characteristics, decision timing, action spaces, and control arrangements shape the choice of method?

RQ4: Where does the evidence fall short of establishing reliable operation in a VPP?

2. Review methodology

The narrative review covers primary research from 2019 to August 15, 2026. IEEE Xplore, Web of Science Core Collection, Scopus, Google Scholar, and CNKI make up the documented search scope; publisher pages and DOI records provide supplementary checks. The keyword themes pair virtual power plants with learning methods and operating tasks such as aggregation, capacity forecasting, bidding, dispatch, and network coordination.

A paper qualified only if it described a VPP explicitly and tied learning or data-driven modeling to an operating decision. This also admitted forecasting and aggregation papers when they identified the scheduling task served by their outputs. All included papers had to be peer-reviewed primary studies reporting at least one quantitative result. Generic microgrid control and standalone prediction were outside the scope. Conceptual discussions, theses, reviews, and publications outside the stated dates were excluded, leaving fifteen journal articles; each study was counted once.

The studies are grouped by operating task, with each result read against the baseline and setting in which it was obtained. A reported runtime, for example, may cover policy training or the completion of a dispatch task; the two are discussed separately and retain their original units. For constraint results, the resource bounds and network model define the choices available to the controller and hence the scope of the reported outcome.

3. Literature review

3.1. Deliverable flexibility

Figure 1 shows how Subramanya et al. connect measurements, forecasting services, and a commercial VPP through a generic application programming interface. The photovoltaic use case estimates day-ahead capacity for reserve planning [1].

DOI: 10.54254/2755-2721/2026.LD37913

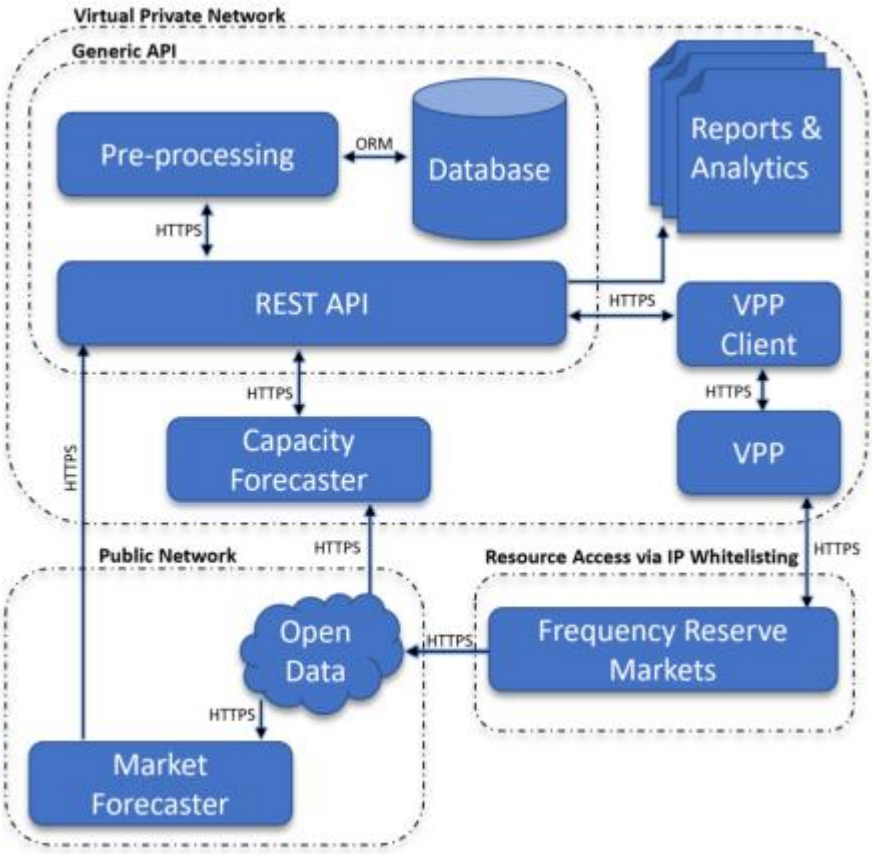

Figure 1. Generic VPP application programming interface architecture

For controllable building demand, Cui et al. use piecewise-linear relationships to identify thermal dynamics and construct virtual-storage models. Their two-stage aggregation updates power bounds with feasible device-level disaggregation in view [2]. These physical bounds do not determine whether customers will participate. Luo et al. classify responses to incentives and use differentiated payments in VPP cost minimization, considering customer benefit alongside procurement cost [3].

Bai et al. use load and dispatch records to describe VPP clusters, including resource reliability. Their prediction comparison reports root mean square error reductions of 44.22% against a long short-term memory baseline and 54.41% against a back-propagation neural network [4].

Forecasting describes the capacity expected to be available, whereas a disaggregation model checks how an aggregate request can be assigned to devices [1, 2, 4]. Customer incentives introduce a further condition: physically feasible capacity is useful only when its owners participate [2, 3].

3.2. Forecasts and market commitments

Sadeghi et al. combine bidirectional long short-term memory forecasts with day-ahead energy and frequency-regulation bidding under a California Independent System Operator market formulation [5]. For November 30, 2018, the profit calculation using forecasts is $4,156.981, compared with $4,319.255 using actual data; the source reports a 3.75% discrepancy [5]. Decision-tree inputs produce $6,309.863 and a much larger discrepancy of 46.08% [5]. The higher calculated profit reflects the effect of forecast inputs on the model. Realized returns would require settlement under common observed market and resource conditions [5].

DOI: 10.54254/2755-2721/2026.LD37913

Zheng et al. embed battery dispatch in forecast training through differentiable optimization, so the cost of the resulting schedule guides the predictor [6]. An error then matters through the charging decision it changes, including whether capacity remains available to reschedule it. The operating objective evaluates a forecast-based schedule in one design and also shapes the predictor during training in the other [5, 6]. Forecast accuracy alone therefore cannot settle their relative value for scheduling.

3.3. Real-time allocation

Cloud training supplies the economic-dispatch policy used by Lin et al.'s edge agents. Their asynchronous actor–critic design is compared with deep deterministic policy gradient [7].

For regulation disaggregation, Yi et al. build adaptation into the transfer from simulator training to operation. Their soft actor–critic policy continues to update in the operating environment using sharpness-aware minimization, with validation carried out in simulation [8].

For residential resources, an allocation made now changes the capacity available later. Deng et al. combine gated recurrent units, constrained soft actor–critic, and Lagrangian relaxation to connect day-ahead and real-time decisions [9]. He et al. study individual task assignments, using a deep Q-network to choose among ten peaking and baseload units. Success is defined by completing the assigned work before its deadline, and response time spans arrival to completion [10].

All four designs learn allocations, but Lin et al. separate cloud training from edge dispatch, while Yi et al. address adaptation after simulator training [7, 8]. Deng et al.'s cross-period schedules and He et al.'s deadline-based assignments differ in both the horizon and the definition of a successful allocation [9, 10].

3.4. Network coordination and constraints

Liu et al. use communication-enhanced twin delayed deep deterministic policy gradient (TD3) for network-constrained VPPs whose pricing decisions interact through internal and external markets [11]. Sun and Lu study a hierarchy in which a distribution-system operator sets transaction prices and VPPs adjust local schedules. They test parameter-sharing TD3 with three VPPs on a modified 33-node network [12]. For this test, Table 1 reports execution times alongside the operator's and VPP1's costs [12].

Table 1. Selected costs and execution times reported by Sun and Lu [12]

Method	Distribution-systemoperator cost ($)	VPP1 cost ($)	Reported executiontime (s)
Parameter-sharing TD3	814.69	206.03	1.79
TD3	852.05	211.45	3.11
Deep deterministic policy gradient	878.17	235.19	2.74
Model-based method	819.43	205.42	112.91

Feng et al.'s VPPs with electric vehicles coordinate through federated gradient exchange [13]. Two rules based on gradient distance filter noisy, negatively scaled, or malicious contributions, while a stochastically controlled stochastic-gradient procedure improves sampling efficiency [13].

DOI: 10.54254/2755-2721/2026.LD37913

An adjustable-space representation connects the local controllers to Xue et al.'s network-level TD3 scheduler [14]. The modified 33-node test gives millisecond-scale decisions and a 1.46% difference from the optimal value, with no violations observed during training or testing; a separate 116-node case examines scalability [14]. For outages, Wei et al. combine fast-response VPP injections with Lagrangian soft actor–critic in preventive and corrective security-constrained optimal power flow, using IEEE 30-bus and 118-bus systems [15].

Gradient filtering operates on the updates exchanged during learning; the adjustable region restricts which choices can pass between network and local controllers [13, 14]. Lagrangian formulations use constraint costs to guide learning, so action-level checks remain necessary even when the expected cumulative cost is bounded [9, 15].

4. Discussion

4.1. Method choice and results

Comparing methods requires identifying what learning replaces. In building aggregation and decision-focused forecasting, the optimizer still selects actions, so the comparison concerns how learning changes the resource description or forecasts inputs [2, 6]. A direct policy changes action selection itself, with the residential formulation also carrying capacity consequences across day-ahead and real-time decisions [9].

Relative to the model-based method in Table 1, parameter-sharing TD3 executes faster and lowers distribution-system operator cost, with a small increase in VPP1's cost [12]. The two methods also use different power-flow models. Sun and Lu attribute the model-based method's voltage violations to the mismatch between its linearized representation and the nonlinear assessment network [12]. A matched network model would help isolate the controller's contribution.

Lin et al.'s 45,000-episode timing result concerns training [7]. It covers a different workload from He et al.'s response time, which runs through task completion [10]. Sun and Lu report training and execution separately, allowing the execution values in Table 1 to be considered against an operating deadline within the tested setting [12].

4.2. Deployment evidence

The aggregation and disaggregation results rely on the resource assumptions used in their respective tests [2, 8, 14]. Their connection can be examined by holding the portfolio, request sequence, and policy fixed, then changing the description of available capacity. Differences in delivery would show how the policy responds to that revised information.

Device measurements from Subramanya et al.'s commercial VPP interfaces could support checks of allocations against actual responses [1]. Yi et al.'s adaptive-disaggregation results come from simulation [8]. A test with missing telemetry or delayed commands would examine how the policy behaves when the information or timing assumed by the controller is disrupted.

Feng et al.'s gradient filters are evaluated against disturbed updates [13]. The confidentiality of federated exchanges remains untested: a separate assessment would need to specify which exchanges an observer can inspect and what local information can be recovered.

DOI: 10.54254/2755-2721/2026.LD37913

5. Conclusion

The value of a learned component lies partly in whether it helps a VPP deliver the promised service. An accepted bid links estimation to control by relying on capacity that must still be available at allocation. If that capacity changes, correct execution of an instruction may no longer be enough to meet the commitment. This makes the consistency between planning assumptions and device operation a central criterion for assessing learning, alongside improvements in the individual component.

For demand-side resources, flexibility describes what the portfolio can offer over the service period, including the consequences of earlier use. A building may meet a request by reducing consumption and still need to recover later, when another request arrives. Customer participation further limits how much of the physically feasible response can be called upon. An assessment confined to the first successful delivery would miss those later restrictions. The capacity carried forward into the remaining schedule therefore has to reflect both the resources already used and the response still available from participants.

The remaining deployment question is whether estimated flexibility survives the transition from planning to delivery as resource conditions change. The comparisons assembled here do not resolve that question for an integrated forecasting and control process. Future evaluation should follow the same request from its capacity estimate through device allocation to measured response. This would help separate an estimate that has become outdated from a failure to execute an otherwise feasible allocation. Repeated observations under changing resource availability would also show when an estimate needs revision and when the portfolio can no longer support the selected service.

References

[1] Subramanya, R., Yli-Ojanperä, M., Sierla, S., Hölttä, T., Valtakari, J., & Vyatkin, V. (2021). A virtual power plant solution for aggregating photovoltaic systems and other distributed energy resources for Northern European primary frequency reserves. *Energies, 14*, Article 1242. https://doi.org/10.3390/en14051242

[2] Cui, X., Liu, S., Ruan, G., & Wang, Y. (2024). Data-driven aggregation of thermal dynamics within building virtual power plants. *Applied Energy, 353*, Article 122126. https://doi.org/10.1016/j.apenergy.2023.122126

[3] Luo, Z., Hong, S., & Ding, Y. (2019). A data mining-driven incentive-based demand response scheme for a virtual power plant. *Applied Energy, 239*, 549–559. https://doi.org/10.1016/j.apenergy.2019.01.142

[4] Bai, X., Fan, Y., Hao, R., & Yu, J. (2025). Data-driven virtual power plant aggregation method. *Electrical Engineering, 107*, 569–578. https://doi.org/10.1007/s00202-024-02544-z

[5] Sadeghi, S., Jahangir, H., Vatandoust, B., Golkar, M. A., Ahmadian, A., & Elkamel, A. (2021). Optimal bidding strategy of a virtual power plant in day-ahead energy and frequency regulation markets: A deep learning-based approach. *International Journal of Electrical Power & Energy Systems, 127*, Article 106646. https://doi.org/10.1016/j.ijepes.2020.106646

[6] Zheng, W., Liu, L., Li, S., Zhang, Z., Zhao, P., Li, T., Zhang, X., & Wang, Z. (2026). Smart predict-then-optimize-based model predictive control for virtual power plants with battery storage. *Journal of Energy Storage, 142*, Article 119606. https://doi.org/10.1016/j.est.2025.119606

[7] Lin, L., Guan, X., Peng, Y., Wang, N., Maharjan, S., & Ohtsuki, T. (2020). Deep reinforcement learning for economic dispatch of virtual power plant in Internet of Energy. *IEEE Internet of Things Journal, 7*, 6288–6301. https://doi.org/10.1109/JIOT.2020.2966232

[8] Yi, Z., Xu, Y., Wang, X., Gu, W., Sun, H., Wu, Q., & Wu, C. (2022). An improved two-stage deep reinforcement learning approach for regulation service disaggregation in a virtual power plant. *IEEE Transactions on Smart Grid, 13*, 2844–2858. https://doi.org/10.1109/TSG.2022.3162828

[9] Deng, X., Chen, Y., Fan, D., Liu, Y., & Ma, C. (2024). GRU-integrated constrained soft actor-critic learning enabled fully distributed scheduling strategy for residential virtual power plant. *Global Energy Interconnection, 7*, 117–129. https://doi.org/10.1016/j.gloei.2024.04.001

DOI: 10.54254/2755-2721/2026.LD37913

[10] He, S., Cui, W., Li, G., Xu, H., Chen, X., & Tai, Y. (2025). Intelligent scheduling of virtual power plants based on deep reinforcement learning. *Computers, Materials & Continua, 84*, 861–886. https://doi.org/10.32604/cmc.2025.063979

[11] Liu, X., Li, S., & Zhu, J. (2023). Optimal coordination for multiple network-constrained VPPs via multi-agent deep reinforcement learning. *IEEE Transactions on Smart Grid, 14*, 3016–3031. https://doi.org/10.1109/TSG.2022.3225814

[12] Sun, Z., & Lu, T. (2024). Collaborative operation optimization of distribution system and virtual power plants using multi-agent deep reinforcement learning with parameter-sharing mechanism. *IET Generation, Transmission & Distribution, 18*, 39–49. https://doi.org/10.1049/gtd2.13037

[13] Feng, B., Liu, Z., Huang, G., & Guo, C. (2023). Robust federated deep reinforcement learning for optimal control in multiple virtual power plants with electric vehicles. *Applied Energy, 349*, Article 121615. https://doi.org/10.1016/j.apenergy.2023.121615

[14] Xue, L., Zhang, Y., Wang, J., Li, H., & Li, F. (2024). Privacy-preserving multi-level co-regulation of VPPs via hierarchical safe deep reinforcement learning. *Applied Energy, 371*, Article 123654. https://doi.org/10.1016/j.apenergy.2024.123654

[15] Wei, X., Chan, K. W., Wang, G., Hu, Z., Zhu, Z., & Zhang, X. (2025). Robust preventive and corrective security-constrained OPF for worst contingencies with the adoption of VPP: A safe reinforcement learning approach. *Applied Energy, 380*, Article 124970. https://doi.org/10.1016/j.apenergy.2024.124970